ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science in the Public Interest
2020, Vol. 21(2) 43–49

DOI: 10.1177/1529100620933847
www.psychologicalscience.org/PSPI
SAGE

Interventions to Do Real-World Good: Generalization and Persistence

C. Shawn Green
Department of Psychology, University of Wisconsin-Madison

In his 1955 address to the National Academy of Sciences, Richard Feynman delineated three key ways in which he saw science as having value (Feynman, 1955). One of these ways was the simple "intellectual enjoyment which some people get from reading and learning and thinking" (p. 13). For many scientists, there is intrinsic value in simply coming to understand how things work. They feel a certain joy when aspects of the world that previously seemed completely mysterious or idiosyncratic become less so. And this is true regardless of how the knowledge is eventually put to use. Yet it is inarguably the case that those eventual uses represent the greatest long-term value of science to our broader society. As Feynman said in discussing this second way that science has value, science is important because it "enables us to do all kinds of things and to make all kinds of things" (p. 13). In other words, increasing scientific understanding of a domain provides for the increasing possibility that we can apply some degree of control in the domain. Science offers the promise that we can manipulate, and thus potentially master, our circumstances.

This core notion certainly permeates the behavioral sciences. Throughout the literature, one consistently sees manifestations of the idea that if we come to truly understand the mechanics by which human abilities, skills, knowledge, and other life outcomes emerge, then we might be able to purposefully intervene so as to alter those outcomes for the better. And although we are absolutely (very, very) far from mastering our circumstances in this domain, there are at least many reasons to be hopeful that such goals will eventually be within our reach. Such reasons for optimism include, for example, promising and ever-growing bodies of research on behavioral interventions meant to increase mental health and well-being (Creswell, 2017; Davidson & Dahl, 2018), interventions meant to decrease bias and prejudicial actions (Lemmer & Wager, 2015; Paluck & Green, 2009), interventions meant to increase cognitive and perceptual functioning (Au et al., 2015; Bediou et al., 2018; Deveau, Jaeggi, Zordan, Phung, & Seitz, 2014), and interventions in the educational sphere, such as those to promote reading abilities (Bus & van IJzendoorn, 1999; Kim & Quinn, 2013).

Yet in considering previous work, as well as in evaluating the potential of future work, it is critical to recognize that in most cases of human behavior, truly "doing good" necessitates that the effects of interventions meet at least two key criteria: (a) The impact of the given intervention needs to generalize reasonably broadly and (b) the impact of the given intervention needs to be enduring. If the impact of an intervention is exceedingly narrow, or if the positive impact lasts for only a short period of time, this will obviously reduce the real-world good that will be realized from the intervention. It is therefore somewhat unfortunate that the field of human learning has consistently run into significant obstacles on both key fronts—generalization and persistence.

Generalization

Starting with generalization, it is first worth noting that it is often quite easy to design interventions that improve individuals' abilities to perform discrete tasks. In fact, given proper experience, humans can sometimes reach seemingly amazing levels of competence on single tasks. For instance, within the world of perceptual interventions, one common training task is known as the Vernier acuity task (Fahle & Edelman, 1993). In this task, two small vertically oriented lines are presented on a computer screen, one on top of the other, but the top line is offset slightly to the right or the left of the bottom line. The participants' task is simply to indicate the direction of this offset. While this task seems quite simple on the surface, given long-term training, human performance on this task can eventually reach astounding levels, with some individuals ultimately being able to determine the direction of offsets

Corresponding Author:
C. Shawn Green, University of Wisconsin-Madison, Department of Psychology, 1202 W. Johnson St., Madison, WI 53706
E-mail: cshawn.green@wisc.edu

that are as small as 1 s of arc (Guinness World Records, 1984). To put that level of performance into real-world terms, if you look at your pinky nail at arm's length and imagine dividing the nail up into 360 even pieces, 1 s of arc would be the width of just one of those pieces. And this result is far from unique. Across basically every research subdomain that examines human learning, the literature is in total consensus—humans are exceptionally good at learning to perform individual tasks via dedicated practice (Ericsson, Krampe, & Tesch-Romer, 1993).

Yet although people are indeed quite good at learning to perform the individual tasks that they are trained on, the learning that emerges from such interventions very commonly fails to transfer to new contexts or tasks. For example, if, after extensive training on the vertical-line version of the Vernier acuity task discussed above, the lines are then rotated by 90° (i.e., the lines are switched to being horizontally aligned with one another with a slight vertical offset between them), performance often falls back entirely to baseline. In other words, none of the learning that occurred for vertical lines transfers to the same lines just rotated by 90° (Fahle, 2005). Similar results have been seen throughout the perceptual-training literature, where performance frequently fails to transfer to stimuli if they are moved to new spatial positions or if the participants are moved slightly closer or further from the stimuli (Fiorentini & Berardi, 1987; Green, Banai, Lu, & Bavelier, 2018; Sagi, 2011; Seitz, 2017). Given that real-world impact in the case of perceptual interventions necessarily requires that improvements extend to all the various types of stimuli that individuals will see in the real world, this degree of specificity is a substantial obstacle.

Critically, this tendency toward specificity of learning is not unique to perceptual training. Similar lack of transfer has been documented in interventions throughout the behavioral sciences, even in those areas, such as education, in which it might be imagined that the knowledge or skills that are being taught would be more generalizable (Barnett & Ceci, 2002). To give one particularly interesting example, research has suggested that when young students are given practice with solving mathematics equations, the problems they are given overwhelmingly take the general form of $X + Y =$ ____, where the operands are all on one side of the equation and are separated from the equal sign by the to-be-filled-in blank (e.g., $4 + 5 =$ ____ ; or $8 - 3 =$ ____; McNeil et al., 2006). Although students absolutely learn to solve equations in this format, many then subsequently struggle to solve more complex problems that do not take the simple form above. For instance, given the equation, $4 + 5 = 3 +$ ____, some students will tend to fill in the blank with "9" (which would be consistent with the incorrect belief that the blank is where they should put the sum of the quantities to the left of the equal sign), whereas other students will put in "12" (consistent with the erroneous belief the blank is where they should put in the sum of the quantities that appear before the blank; Perry, Church, & Goldin-Meadow, 1988). This is thus an excellent example of a case in which generalization fails because the knowledge that was extracted from the initial training was overly narrow. It is noteworthy that in the case of mathematics, the failure to extract more generalizable forms of knowledge from earlier learning then subsequently predicts a number of learning failures further down the line—such as issues with transitions to algebra (Knuth, Stephens, McNeil, & Alibali, 2006). In all, the tendency toward learning specificity is so prevalent across the behavioral sciences, and runs so directly counter to the goal of producing real-world good, that it has been dubbed the "curse of specificity" (Bavelier, Green, Pouget, & Schrater, 2012; Deveau & Seitz, 2014; Green & Bavelier, 2008).

Persistence

Next, with regard to producing positive impact that persists through time, scientists again appear to be working directly against what is probably the prevailing human tendency. Some of the issues to be faced when attempting to design interventions for persistent impact reflect purely internal mechanics. For example, across basically every subdomain of behavioral science, there is the shared recognition that learned performance will tend to diminish through time (at least in the absence of additional intervention). Depending on the subdomain, this might be referred to as "forgetting" (Ebbinghaus, 1885; Murre & Dros, 2015), "skill decay" (Arthur, Bennett, Stanush, & McNelly, 1998), or some other similar phrase, but each of these terms reflects the fundamental fact that the gains that are acquired from an intervention very frequently weaken through time (Davis & Zhong, 2017).

Furthermore, in addition to the various internal mechanisms that will tend to drive learned improvements back toward baseline, there are a host of other issues that will also fight against measured persistence of impact. These include issues that arise because impact is nearly always calculated as a *relative change* in outcome between those who did and those did not receive an intervention (Green, Strobach, & Schubert, 2014). Consequently, even if the skills that are developed via an intervention do persist, the net positive real-world impact will nonetheless still be reduced through time if those individuals who did not receive the intervention also eventually develop the skills of interest. Indeed, in many learning domains, humans show a roughly exponential learning curve (i.e., where they learn a constant percentage of what remains to be learned per unit time; Dosher & Lu, 2007). In absolute terms, this functional form will naturally produce the tendency for individuals who are initially behind to "catch up" given similar experience.

For example, consider a hypothetical domain in which there is a maximum of 10 "units" of some quantity to learn (e.g., entirely for the purpose of discourse, imagine dividing reading ability into 10 discrete levels of performance). In this hypothetical domain, Group 1 receives an intervention through which they learn 4 units of the skill (e.g., some prekindergarten reading intervention), whereas Group 2 receives no intervention. After the intervention, the two groups then proceed to have similar experiences, in particular going to school where reading skills are taught. Group 1 will start school at 4 units of skill, whereas Group 2 will start at 0 units of skill. And then, going forward, at each time point, each group learns 50% of what is left for them to learn. Over three time steps, Group 1 will go from 4 to 7 (i.e., moving up by 3 units represents an increase of 50% of what was left to learn), then to 8.5, and then to 9.25. Over the same three time steps, Group 2 will go from 0 to 5, to 7.5, and to 8.75. What was a difference of 4 units directly after the intervention very quickly becomes a difference of only 0.5, an eight-fold decrease in the absolute magnitude of the impact. And although there are reasons why reading skills might not be a perfect analogy here, it remains the case that, broadly speaking, the basics of how humans learn will frequently produce catch-up outcomes.

Together, then, all of the issues discussed above will contribute to the global phenomenon of "fade-out," defined by Bailey, Duncan, Cunha, Foorman, and Yeager (2020) in their current *PSPI* piece (this issue), as a "temporal pattern of diminishing effect sizes following the end of an effective intervention" (p. 60). In examining this phenomenon primarily in the domain of educational interventions, the authors first and foremost summarize the evidence that fade-out in educational interventions is exceedingly widespread and thus of clear concern for public policy. Indeed, each of several meta-analyses reviewed by Bailey et al. (2020), which together span a broad range of education-related targets, provided clear evidence for fade-out. For example, a meta-analysis by Hattie and colleagues examined the impact of 51 different study-skill interventions (e.g., interventions meant to improve cognitive strategies for learning, such as those involved in note-taking; Hattie, Biggs, & Purdie, 1996). This meta-analysis found a beneficial effect of study-skill interventions in the medium range when the impact was assessed directly after the end of the interventions. Yet in the 30 interventions that also included a long-term follow-up, the magnitude of the beneficial effect fell by around 75% in these later assessments. It is noteworthy that this same meta-analysis also explicitly noted that when the goal of those interventions was to produce meaningful transfer, the interventions tended to have the smallest effect sizes from the outset, which is again indicative of the difficulty in hitting both the generalization and persistence criteria.

Need for Long-Term Follow-Up

After putting forward a convincing case that fade-out is, in fact, a pervasive outcome of educational interventions (and one that cannot be fully explained by simple artefacts of methodology or publication practices), Bailey and colleagues (2020) go on to make a host of other key points that will be invaluable both for scientists designing and testing the impact of interventions, as well as for policymakers considering whether to implement interventions in the real-world. Of these points, first and foremost is the need to more deeply consider the impact of interventions as long-term trajectories. Unfortunately, such a consideration is, at least to date, a relative rarity in the scientific literature. Indeed, the most common intervention design across the behavioral sciences is a simple pretest → intervention → posttest design (Green et al., 2019). Under this design, participants first complete a pretest assessment consisting of a small set of outcome measures. The participants are then randomly assigned to complete one of two conditions (usually an active intervention or a control intervention). Finally, after the end of the intervention, the participants are posttested on the same basic measures as were taken at pretest. The critical question is then whether the group that received the active intervention improved more from pretest to posttest than those who were in the control group.

Several of the meta-analyses reviewed by Bailey and colleagues (2020) speak directly to the uncommonness of research studies that also include longer term follow-up assessments. For example, as discussed above, in the meta-analysis by Hattie and colleagues (1996), only around 60% of the interventions included an additional follow-up assessment; of those that did, the average follow-up occurred only a few months after the intervention. Likewise, in a meta-analysis by Takacs and Kassai (2019) that examined the impact of interventions meant to enhance executive functions, only about 15% of studies included a follow-up at all; of those that did, the follow-up again largely came within a few months of the end of the intervention. Simply put, long-term follow-ups to behavioral interventions are currently the exception rather than the rule. This is an unfortunate state of affairs, given that there is obviously no way to assess whether fade-out is observed without such long-term follow-ups.

Isolating Mechanisms of Fade-Out

A second key point made by Bailey and colleagues (2020) is that although fade-out is a critical outcome to look for, and one that can necessarily be observed only when viewing the impact of an intervention as a trajectory, there are a number of fundamentally different mechanisms that can potentially produce a fade-out result. As discussed above, some of these mechanisms will reflect a failure of the intervention to generalize beyond the confines of the intervention itself. This type of mechanism may be especially prevalent in designs in which the initial measures are exceedingly similar to the training provided during the intervention (i.e., "teaching to the test"), but later measures are considerably more dissimilar to the intervention. The observation of fade-out in this situation could reflect a failure of generalization that was present from the outset, rather than a decrease through time per se.

Consider another hypothetical scenario: Investigators find that some Outcome 1 (that is very similar to the intervention) was positively affected when assessed directly after the end of the intervention, but then they find that Outcome 2 (that is far less similar to the intervention) was not affected by the intervention when assessed 2 years later. In this situation it is entirely possible that this pattern of results does not reflect a reduction in the impact of the intervention over the period of 2 years; rather, the intervention might never have actually generalized to Outcome 2 in the first place. In some research domains, this type of "failure to generalize" mechanism can potentially be detected from the outset, simply by increasing the number of outcome measures taken at pretest and posttest. In particular, this speaks directly to the importance of always including a number of measures that are core to the construct(s) of interest while simultaneously being as far from the exact experiences of the intervention as possible (Green et al., 2019). Yet there are situations in which assessing whether this failure to generalize mechanism is at play can be exceptionally difficult. This is particularly true in domains, such as many educational interventions, in which the types of behaviors of interest change through time via some developmental trajectory (e.g., one cannot reasonably add algebra questions to the posttest for a first-grade intervention focused on addition and subtraction; the impact of the intervention on algebra performance can necessarily only be tested down the line).

Other mechanisms will produce a fade-out outcome but they do not reflect a failure of generalization; rather, they reflect a true reduction through time. The presence of forgetting or skill decay can potentially be detected simply by making the same measurement repeatedly through time. Here, though, another important caveat needs to be considered. Specifically, it is crucial to recognize that repeated testing can act as a form of training in and of itself. If, for example, a participant is asked to complete a particular outcome measure four times a year for 3 years, it's very likely that they'll show significant improvements on that measure simply because of increasing familiarity with the test. This test-based learning will in turn make it progressively more difficult to isolate the impact of the intervention, particularly given the fact that the impact of the intervention will undoubtedly be assessed relative to a control group that is also being repeatedly tested (and thus will also be learning from the repeated testing; for a more thorough discussion of this issue, see Green et al., 2014).

Although delineating the mechanism or mechanisms that are responsible for the fade-out of an intervention is clearly difficult, doing so is critical because this will guide the process of altering future interventions so as to potentially reduce or eliminate the presence of fade-out. Again, thinking in terms of the trajectory of impact is crucial to this endeavor; producing persistence requires identifying what types of changes will be needed to see positive long-term outcomes, not just increased performance on whatever measures are taken directly at the end of the intervention. Speaking directly to this point, Bailey and colleagues (2020) discuss the need to focus on what they dub "trifecta" skills. These are skills that (a) are malleable (i.e., can be altered via experience); (b) are fundamental (i.e., underlie a wide variety of outcomes); and (c) require the presence of an intervention to develop (i.e., skills that will already be the focus of normal schooling might not be good options because this will greatly increase the chances of "catchup").

Similar lines of thinking underpin theory in a variety of behavioral intervention domains outside of education. For instance, these same basic essential criteria guide the selection of targets in the cognitive-training literature (Deveau et al., 2014; Nahum, Lee, & Merzenich, 2013). Indeed, the large amount of research on training executive functions, attention, and fluid intelligence have all arisen, at least to some degree, because these functions meet Criteria 2 above (i.e., they are fundamental). Each of these cognitive functions predicts an enormous variety of real-world behavioral outcomes, from high school test scores to job-performance and income levels (Green & Newcombe, 2020; Karbach & Unger, 2014; Karbach & Verhaeghen, 2014). The major difficulty in these fields has been assessing whether Criteria 1 is actually met (i.e., the skills are malleable). It is clearly the case that individuals can improve on tasks that have been designed to measure executive function, attention, and/or fluid intelligence. Yet which patterns of improvement across tasks meant to tap the same core construct are indicative of true change in the construct is still a matter of considerable debate in the field (e.g., how many different working memory tasks would one need to show improvement on for it to be convincing that "working memory" was truly improved? (Au et al., 2015; Melby-Lervåg, Redick, & Hulme, 2016; Redick, Shipstead, Wiemers, Melby-Lervag, & Hulme, 2015).

Promoting Larger Long-Term Effects

A final way that thinking in terms of trajectories will inform designs for persistence, is that it becomes absolutely clear how incredibly tiny almost any possible intervention (that a scientist might be able to reasonably test) is relative to individuals' full life experiences. In the cognitive training domain for instance, although there are some individual examples of interventions that last for 50 hr or more (Green, Pouget, & Bavelier, 2010; Schmiedek, Lövdén, & Lindenberger, 2010), most are far shorter (Bediou et al., 2018). And 50 hr represents only a bit more than 2 days in a lifetime. The question thus becomes, "Can an intervention cause participants to, in some way, alter their future experiences in a positive manner?" Although the scale of an intervention is likely to be able to produce only a tiny deflection in trajectory, if that tiny deflection fundamentally alters the path that an individual takes, the total impact of the intervention can be far broader. For example, in older adults, an intervention that produces an improvement in the ability to hear speech in noise could massively affect the long-term trajectory that the individuals follow (Pichora-Fuller, Mick, & Reed, 2015). Better hearing is associated with more interest in social interactions. Those social interactions could in turn involve a great deal of additional mental and physical stimulation, both of which are key for healthy cognitive aging. In the end, the most proximal cause of the healthy aging would be this mental and physical stimulation, but such stimulation would have been caused by the initial hearing intervention. Bailey and colleagues (2020) discuss this primarily in the form of "institutional gateways," of which there are many in the educational sphere. And although the overall evidence is somewhat mixed, this manner of thinking is likely to be key, given the goal of producing real-world, long-term impact.

Future Implications

Finally, in turning to the future implications, we come to Feynman's third value of science—that "the scientist has a lot of experience with ignorance and doubt and uncertainty . . . we have found it of paramount importance that to progress, we must recognize the ignorance and leave room for doubt" (p. 14). In essence, it is important for us to keep in mind what things we do and do not know. And in the case of fade-out, there are many open questions. Consequently, the suggestions by Bailey and colleagues (2020) are extremely well aligned both with the areas of most substantial uncertainty and with how to fill the various gaps. Although we have compelling evidence that fade-out is an issue to be taken seriously, our understanding of the mechanisms that underlie fade-out is necessarily more uncertain. Thus, as scientists, it is important for us to take methodological steps to reduce that uncertainty (e.g., more regularly including long-term follow-up assessments, utilizing broader assessment batteries to determine whether an intervention has produced generalizable changes in function). Such steps, though, may require concomitant shifts in funding sources, in terms of both absolute funding levels (implementing these changes can increase costs substantially) and persistence of funding through time (Green et al., 2019). Critically, not only will these shifts increase the rate at which scientific knowledge grows, they will also result in far more informed policy. For instance, the failure to consider lots of possible outcome measures may mask true impact caused by an intervention. If, for example, a mathematics intervention produces no change in standardized test scores relative to a control group, but it does produce a reduction in the need for teaching-assistant time (i.e., the students hit the same level of performance, but required fewer resources to do so), this may very well represent meaningful impact.

In the end, as Feynman put it, "It is not unreasonable that we grapple with problems. Our responsibility is to do what we can, learn what we can, improve the solutions and pass them on" (pp. 15).

Transparency

Editor: Nora Newcombe

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

Funding was provided in part by Grant N00014-17-1-2049 from the Office of Naval Research (to C. S. Green).

References

Arthur, W., Jr., Bennett, W., Jr., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance*, *11*, 57–101.

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *22*, 366–377. doi:10.3758/s13423-014-0699-x

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, *21*, 55–97. doi:10.1177/1529100620915848

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.

Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience*, *35*, 391–416.

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, *144*, 77–110. doi:10.1037/bul0000130

Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, *91*, 403–414.

Creswell, J. D. (2017). Mindfulness interventions. *Annual Review of Psychology*, *68*, 491–516. doi:10.1146/annurev-psych-042716-051139

Davidson, R. J., & Dahl, C. J. (2018). Outstanding challenges in scientific research on mindfulness and meditation. *Perspectives on Psychological Science*, *13*, 62–65. doi:10.1177/1745691617718358

Davis, R. L., & Zhong, Y. (2017). The biology of forgetting: A perspective. *Neuron*, *95*, 490–503. doi:10.1016/j.neuron.2017.05.039

Deveau, J., Jaeggi, S. M., Zordan, V., Phung, C., & Seitz, A. R. (2014). How to build better memory training games. *Frontiers in Systems Neuroscience*, *8*, Article 243. doi:10.3389/fnsys.2014.00243

Deveau, J., & Seitz, A. R. (2014). Applying perceptual learning to achieve practical changes in vision. *Frontiers in Psychology*, *5*, Article 1166. doi:10.3389/fpsyg.2014.01166

Dosher, B. A., & Lu, Z. (2007). The functional form of performance improvements in perceptual learning: Learning rates and transfer. *Psychological Science*, *18*, 531–539. doi:10.1111/j.1467-9280.2007.01934.x

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie* [About memory: research on experimental psychology]. Leipzig, Germany: Duncker & Humblot.

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

Fahle, M. (2005). Perceptual learning: Specificity versus generalization. *Current Opinion in Neurobiology*, *15*, 154–160.

Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research*, *33*, 397–412.

Feynman, R. P. (1955). The value of science. *Engineering and Science*, *19*, 13–15.

Fiorentini, A., & Berardi, N. (1987). Interhemispheric transfer of visual information in humans: Spatial characteristics. *Journal of Physiology*, *384*, 633–647.

Green, C. S., Banai, K., Lu, Z.-L., & Bavelier, D. (2018). Perceptual learning. In J. Serences & J. T. Wixted (Eds.), *Steven's handbook of experimental psychology. Volume 2: Sensation, perception & attention* (pp. 755–802). New York, NY: John Wiley.

Green, C. S., & Bavelier, D. (2008). Exercising your brain: A review of human brain plasticity and training-induced learning. *Psychology and Aging*, *23*, 692–701.

Green, C. S., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge, U., Ball, K. K., & Witt, C. M. (2019). Improving methodological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Enhancement*, *3*, 2–29.

Green, C. S., & Newcombe, N. S. (2020). Cognitive training: How evidence, controversies, and challenges inform education policy. *Policy Insights from the Behavioral and Brain Sciences*, *7*, 80–86.

Green, C. S., Pouget, A., & Bavelier, D. (2010). Improved probabilistic inference as a general mechanism for learning with action video games. *Current Biology*, *23*, 1573–1579.

Green, C. S., Strobach, T., & Schubert, T. (2014). On methodological standards in training and transfer experiments. *Psychological Research*, *78*, 756–772. doi:10.1007/s00426-013-0535-3

Guinness World Records. (1984). *Highest hyperacuity*. Retrieved from https://www.guinnessworldrecords.com/world-records/highest-hyperacuity/

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, *66*, 99–136.

Karbach, J., & Unger, K. (2014). Executive control training from middle childhood to adolescence. *Frontiers in Psychology*, *5*, Article 390. doi:10.3389/fpsyg.2014.00390

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. doi:10.1177/0956797614548725

Kim, J. S., & Quinn, D. M. (2013). The effects of summer reading on low-income children's literacy achievement from Kindergarten to Grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research*, *83*, 386–431.

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education*, *37*, 297–312.

Lemmer, G., & Wager, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology*, *45*, 152–168.

McNeil, N. M., Grandau, L., Knuth, E. J., Alibali, M. W., Stephens, A. C., Hattikudur, S., & Krill, D. E. (2006). Middle-school students' understanding of the equal sign: The books they read can't help. *Cognition and Instruction*, *24*, 367–385.

Melby-Lervåg, M., Redick, T. S., & 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 of Psychological Science*, *11*, 512–534.

Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLOS ONE*, *10*(7), Article e0120644. doi:10.1371/journal.pone.0120644

Nahum, M., Lee, H., & Merzenich, M. M. (2013). Principles of neuroplasticity-based rehabilitation. *Progress in Brain Research*, *207*, 141–171. doi:10.1016/B978-0-444-63327-9.00009-6

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology*, *60*, 339–367.

Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, *3*, 359–400.

Pichora-Fuller, M. K., Mick, P., & Reed, M. (2015). Hearing, cognition, and healthy aging: Social and public health implications of the links between age-related declines in hearing and cognition. *Seminars in Hearing*, *36*, 122–139. doi:10.1055/s-0035-1555116

Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-Lervag, M., & Hulme, C. (2015). What's working in working memory training? An educational perspective. *Educational Psychology Review*, *27*, 617–633. doi:10.1007/s10648-015-9314-6

Sagi, D. (2011). Perceptual learning in *Vision Research*. *Vision Research*, *51*, 1552–1566. doi:10.1016/j.visres.2010.10.019

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience*, *2*, Article 27. doi:10.3389/fnagi.2010.00027

Seitz, A. R. (2017). Perceptual learning. *Current Biology*, *27*, R631–R636. doi:10.1016/j.cub.2017.05.053

Takacs, Z. K., & Kassai, R. (2019). The efficacy of different interventions to foster children's executive function skills: A series of meta-analyses. *Psychological Bulletin*, *145*, 653–697.