Published in final edited form as:

J Cogn Enhanc. 2018 March ; 2(1): 78–87. doi:10.1007/s41465-017-0058-8.

A new framework of design and continuous evaluation to improve brain training

Aaron R. Seitz

University of California, Riverside

Abstract

Effective teaching critically relies upon effective evaluation because evaluation is required to gain understanding of existing abilities and, in turn, determine learning outcomes. Methods of effective evaluation are surprisingly elusive in many fields and this limits our understanding of which training methods are truly effective. In the present article, issues of effective evaluation are discussed in the context of “brain training,” an exciting but much criticized field. Problems in test validity in the field of brain training, parallel those in many other fields; such as deficiencies in test reliability, teaching to the test, expectation effects, as well as statistical rigor. Here we review these issues and discuss how commonalities between the goals of evaluation and adaptive training procedures suggest a new paradigm that synthesizes evaluation and training. We suggest that continuous evaluation, where testing is integrated into the training, may provide a path towards greater reliability of skill evaluation, through longitudinal sampling, and validity, through better alignment of evaluation activities in respect to learning objectives.

Introduction

Teaching and testing necessarily go hand in hand; how else can we understand whether learning objectives are achieved? However, the challenge of establishing reliable and valid tests is an ubiquitous problem. For example, in school settings a common criticism is that the rise of standardized testing has been accompanied by problems of “teaching to the test” (Volante, 2004). The essential problem is that if students, teachers and schools are all evaluated by test performance then the goal of education should be to ensure that students perform well on these tests. This raises the question of *test validity* (the extent to which good performance on the test relates to actual mastery of learning objectives). For example, if performance on these tests was perfectly correlated with student knowledge and eventual ability to use this knowledge effectively in the workplace, then it would make perfect sense to ensure that everyone aced these exams as it would likely be the case that training for the exams would lead to the knowledge assessed and predicted performance. However, the reality is that most tests, at best, evaluate only a small proportion of the true learning objectives. Further there is an issue of *test reliability* (the extent to which test scores are stable to repeated testing). Tests can be stressful, confusing, and tiresome, among other problems, and thus the same person with the same knowledge might get two very different scores if tested on the same exam at two different times. While problems of test validity and reliability are well-established, ideal solutions are few and far between. Here, we review issues of testing in the context of the field of “Brain Training.” While knowledge, strategy, and skill based learning certainly involve the brain, here we concentrate on training

approaches targeted to improve basic brain processes (vision, hearing, attention/cognitive control, memory, executive function) that have been the focus of typical approaches under the nomenclature of brain training. In light of the particular issues facing brain training we propose a novel method of continuous evaluation within training contexts aimed at achieving more valid and reliable evaluations of learning outcomes.

Of note, while the present discussion of training and evaluation is largely concentrated to the field of brain training, other disciplines face similar issues. In particular, any form of computerized evaluation or training generates substantial performance data that, if properly exploited, can improve learning outcomes. For example, when a student completes a paper based assignment, the record of that student's work is limited to the marks that are left on the page. On the other hand, with a computer assessment, every action of the student can be recorded and time-stamped and potentially used to evaluate that student's strengths and weaknesses in performing the task. Further, computer-based lessons can be changed in real time in response to these evaluations. As such, we are seeing rapid progress in intelligent tutors (Diziol et al., 2010; O'Rourke et al., 2015) and stealth testing (J. V. Shute et al., 2016; V. Shute et al., 2017) where evaluation is embedded in games or training materials. These have been adopted fairly widely in math education and increasingly so in other fields, such as methods used by groups such as Khan Academy or games like MathGarden. However, the field of brain training is relatively young, conventional approaches to assessment are heavily criticized and more effective approaches to testing are very much needed to advance the field.

In this article, we first provide a review of challenges to the field of brain training and then suggest a series of novel approaches to this field, although less novel to others, such as clearly defining training objectives and continuously evaluating performance in respect to these learning objectives can provide a route both to better assessment but also can potentially pave the way to creating more effective brain training methodologies.

Brain Training

The field of brain training is one of substantial excitement and potential, for example who wouldn't want improved perceptual abilities, memory, or clarity of thought? While this may sound like a fantasy, there is support for the premise that training can alter fundamental brain processes and in turn benefit the diversity of tasks that people conduct in daily life that rely upon those abilities (Anguera et al., 2013; Bul et al., 2015; Deveau, Ozer, et al., 2014; Jaeggi et al., 2008; Klingberg et al., 2002; Merzenich et al., 1996; Whitton et al., 2014). For example, if one could improve working memory (our ability to mentally manipulate multiple pieces of information), this could positively impact on our ability to make mathematical calculations, understand complex ideas, or make good decision; all of which rely upon manipulations of information in our working memory. Numerous studies in support of this premise show that perceptual (Deveau, Lovcik, et al., 2014; Polat et al., 2004; Whitton et al., 2014), motor (Hikosaka et al., 2002), and cognitive abilities such as attention (Green et al., 2003), memory (Deveau et al., 2015; Jaeggi et al., 2008; Klingberg et al., 2002) and executive function (Anguera et al., 2013; Dahlin et al., 2008) can be strengthened through training.

This has given rise to an ambitious field of cognitive training with “brain training games” showing promise in ameliorating effects of ADHD (Dovis et al., 2015; Klingberg et al., 2005), Age-Related Cognitive Decline (Smith et al., 2009), Dyslexia (Franceschini et al., 2013), Schizophrenia (Popov et al., 2011), Sensory Deficits (Polat et al., 2012), Brain Injury (Huxlin et al., 2009), in promoting mental fitness in normally functioning individuals (Smith et al., 2009), and even improving skills in high performers such as athletes (Deveau, Ozer, et al., 2014).

Limitations of Brain Training

The effectiveness of existing brain training approaches in producing reliable positive impact upon real world cognition is controversial. One of the most central concerns is that many brain training approaches “teach to the test” by training on cognitive tasks that resemble those used to test cognitive function (Hardy et al., 2015; Klingberg et al., 2005). This has led to criticism of studies that claim positive results (Boot et al., 2011; Owen et al., 2010; Shipstead et al., 2012; Simons et al., 2016). At issue is a lack of agreement on, and consistent use of, methods of evaluation that are reliable, valid, and generalizable to real world cognition.

Accordingly, a number of groups have published recommendations of the study designs to evaluate whether brain training “works.” For example, a recent paper by Simon et al., (2016) suggested that there are no compelling examples effective brain training and urged the use methods of clinical medical studies such as; **Pre-registration** – publishing your study design in advance of the research, **active control conditions** – test the intervention against another intervention that will provide a similar experience and similar expectations in participants but that the researcher expects to lead to different outcomes, **blinding** – to the extent possible ensure that participants and researchers don’t know which condition each participant is in, **randomization** – randomly assign participants between experimental conditions, **large sample sizes** – utilizing a large number of participants can help achieve robustness of results and allow for better statistical evaluations of moderating and mediating factors, **outcome measures of known reliability**– otherwise unreliable outcome measures can lead to spurious results, and **multiple outcome measures** – employing multiple tests that target a given core process allows for independent evaluations of outcome. In addition, they recommend studying dosing effects, using appropriate statistics that control for multiple comparisons and also to publish results of the full study together so that all the data can be reviewed together.

While these recommendations sound reasonable from the perspective of medical testing, they are rather inflexible, inefficient, and overlook the core problem that well agreed upon outcome measures of “known reliability” for real world cognition do not exist and that validity is as important as reliability. In addition, the clinical trial model of research focuses on efficacy to a greater extent than discovery. As a result, forcing every study into a large-scale clinical trial model with pre-registration will discourage researchers from tuning training approaches as they realize that improvements can be made.

Evidence Based Design can inform Brain Training

To overcome limitations of current and proposed evaluation approaches for brain training, it is necessary to better understand the learning objectives of a given training approach, where the approach may fall short, and how to augment that approach to incorporate the elements that are lacking. This is not unlike what we consider appropriate for classroom teaching; effective teaching requires a continuous cycle of evaluating students' knowledge, providing explanations appropriate to student's current knowledge, and then testing whether students can apply the lesson to an example that differs from what was used in the lesson. Then, based upon the results, the teacher will need to provide repeated lessons and evaluations in the area of weakness, or as the students masters the individual lesson add on additional material appropriate to teach the next piece of the larger topic. Evaluation is an integral component of teaching which ensures that the students know the past lessons sufficiently well to successfully learn from the current lesson. If properly quantified classroom teaching can provide a continuous record of students' current knowledge of the topic and their ability to generalize knowledge to novel examples. This model has been embraced by computer based teaching tools, some with better success with others. For example in the case of teaching algebra, there is substantial success with online programs that can effectively analyze student error patterns and adapt lessons to the individual needs of students (Diziol et al., 2010; O'Rourke et al., 2015). A growing field of evidence based design provides a framework in education testing to efficiently and robustly measure student proficiencies (Mislevy, 2013).

Many elements of the evidence based design approach is similar to the adaptive procedures used in brain training; where difficulty is increased following accurate responses and reduced following inaccurate responses in order to determine a "threshold" level of performance where challenge is personalized to one's abilities. In the sections that follow, we give additional background on brain training approaches, address issues of how best to evaluate desirable learning outcomes in this field, and flush out how this model of continuous evaluation can be applied in a robust, reliable and valid manner.

Learning Objectives in Brain Training

Fundamental to any form of instruction is identification of, and alignment with, learning objectives. In the case of brain training, learning objectives are typically "domain general" improvements in various cognitive functions. For example, in the case of working memory, the learning objective is to improve execution of any task for which performance is inhibited by participants' existing working memory capacity. A target example would be the case of a participant who can add two digit numbers on paper, but not mentally. In this case, the participant has full knowledge of the steps required to calculate the correct answer but cannot keep track of and manipulate, the numbers in working memory. Success of a working memory training program for such a person would be to train working memory skills in a non-math context and achieve an outcome of improvement that transfers to the math context (Titz et al., 2014). Other cases might include training hearing skills that transfer to better understanding of speech in noisy environments (Whitton et al., 2014), vision skills that transfer to reading or sports performance (Deveau, Ozer, et al., 2014; Deveau & Seitz, 2014)

and executive function skills that transfer to improved emotional regulation (Hofmann et al., 2012). These examples emphasize the holy grail of brain training; namely to transfer far outside of the training context to benefit real world cognition.

Brain Training can be Highly Specific to the Trained Stimuli

While transfer to real world cognition is the objective, many training paradigms fall short with training benefits that fail to manifest themselves outside of the training task. Much of this can be understood by way of example of the field of Perceptual Learning. Perceptual Learning refers to long-lasting improvement in perceptual abilities as a result of experience (Fahle et al., 2002). A canonical example of perceptual learning is how with extensive experience radiologists learn to see subtle patterns of x-rays that are largely invisible to untrained observers (A. Seitz, 2017).

The field of perceptual learning is one of interest due to the fact that training on visual perception can be highly specific to the trained visual features and can give clues into the stages of processing at which learning occurs. For example, a series of studies conducted by Schoups and colleagues (Schoups et al., 2001; Schoups et al., 1995) showed that training subjects (human and monkey) to discriminate a striped orientation pattern around a particular reference axis yielded learning effects that failed to transfer to the same stimulus when presented to different part of the retina or when rotated around a different axis of orientation. They postulated that these learning effects were consistent with a change of function in the primary visual cortex of the brain, in which individual neurons show a high degree preference to specific line orientations and regions of retinal stimulation. Follow-up physiological studies by this group confirmed these predictions with the demonstration plasticity of orientation tuning in the early visual processing areas of the brain (Raiguel et al., 2006; A. Schoups et al., 2001). Behavioral studies show perceptual learning can be highly specific to a wide range of trained stimulus features including retinotopic location (Crist et al., 1997; Watanabe et al., 2002), line or pattern orientation (Fiorentini et al., 1980; Schoups et al., 2001), direction of motion (Ball et al., 1981; Watanabe et al., 2002), among others. Further, the effects of perceptual learning have been shown to last months, even years (Ball et al., 1981; Crist et al., 2001; Sagi et al., 1994). The exquisite specificity of perceptual learning to the attributes of the trained stimuli and task (Fahle, 2005) have been a barrier to translational relevance of perceptual learning (Deveau & Seitz, 2014).

While the field of perceptual learning represents an extreme case of specific learning, this curse of specificity haunts many fields. Over-specificity is a typical stage in development, for example when first learning the word “dog” children often associate to an individual animal (Piaget, 1952). Although adults are better generalizers, it is still a common occurrence for expertise to be domain specific. For example, radiologists are experts at finding visual defects in medical images but are no better at seeing a subtle picture of a gorilla in an x-ray than untrained subjects (Drew et al., 2013). Documented failures to train domain general skills harken back at least to Thorndike (1913), who extensively reviewed the training literature of that time and found that when training math, language, athletics skills, or other knowledge bases, learning was typically confined to the lesson being taught and failed to generalize outside the training context. One solution is to train on each task that

is required of an individual, which is perhaps why new releases of Microsoft Office are a boon to software training companies. However, finding ways to teach knowledge that generalizes is the target of most educational endeavors and can yield substantial benefits to society.

Mechanisms Governing Specificity or Generalization of Learning

Fortunately, recent studies show how a diversity of factors can contribute to overcoming outcomes of specificity (Hung et al., 2014). For example, in the context of perceptual learning the amount of training (Aberg et al., 2009; Jeter et al., 2010), the difficulty/precision of the stimulus judgments training (Ahissar et al., 1997; Hung et al., 2014) or testing (Jeter et al., 2009), and interleaving different tasks (Szpiro et al., 2014) all give rise to broader transfer profiles. As an approach to overcome learning specificity, Seitz and colleagues created an integrative framework using multiple perceptual learning approaches, including training with a diverse set of stimuli (Xiao et al., 2008), optimized stimulus presentation (Beste et al., 2011), multisensory facilitation (Shams et al., 2008), and consistent reinforcement of training stimuli (Seitz et al., 2009) which have individually contributed to increasing the speed (Seitz et al., 2006), magnitude (Seitz et al., 2006; Vlahou et al., 2012), and generality of learning (Green et al., 2007; Xiao et al., 2008), and incorporated these design principles into a simple video game.

This visual perceptual learning game was essentially a game of whack-a-mole where players clicked on fuzzy blobs (called Gabor filters (Deveau & Seitz, 2014) where stimulus parameters (such as visual intensity, rate of presentation, and similarity with non-target image) were carefully controlled via adaptive procedures that ensured difficulty for each stimulus dimension was personalized to the participant. In this way at each stage of training the program estimated each participants' visual ability and used this knowledge to present consistently difficult task targets even as participants improved on the training task. Importantly these Gabor filters were chosen because neurons in early visual cortex are well-tuned to these stimuli and are considered to be basis function (a set of filters that in combination can represent any stimulus in a stimulus space; for example any visual stimulus can be described mathematically as the sum of Gabor filters of various spatial frequencies, orientations and spatial locations; see below for a more detailed description of basis functions in cognitive training) (Deveau, Lovcik, et al., 2014).

Results of training participants on this game for ~30 25-minute sessions over 2 months showed transfer to real world cognition, including improved reading speed (Deveau & Seitz, 2014), and even on-field performance in collegiate baseball (Deveau, Ozer, et al., 2014). Other researchers have employed similar perceptual learning approaches to benefit individuals with amblyopia (colloquially known as "lazy eye;" (J. Li et al., 2013; R. W. Li et al., 2011), ameliorate effects of peripheral vision loss (Chung, 2011), presbyopia (Polat, 2009), macular degeneration (Baker et al., 2005), vision loss after stroke (Das et al., 2014; Huxlin et al., 2009), and aid other individuals with impaired vision (Huang et al., 2008; Zhou et al., 2012). Similarly, there is an increasing number of studies that indicate well-designed training approaches *can* give rise to real world cognitive benefits including improved understanding of speech in noise (Whitton et al., 2014), ability to multitask

(Anguera et al., 2013), to effectively inhibit irrelevant information (Mishra et al., 2014), improved working memory skills (Au et al., 2015), improved planning and organization skills (Bul et al., 2016). While substantial work is still required to canonize the design principles that can reliably lead to generalizable learning, the above studies provide some promise that generalizable learning is achievable.

The Concept of a Basis Function in Brain Training

The key observation is that achieving learning objectives requires understanding how people learn, specification of the learning objectives, and principled design. For example, take the case of training working memory: if training to remember different sets of colors led to an improvement in the underlying process of working memory, this could manifest as an improvement in performance on all tasks involving working memory, including tests of language, mathematics, intelligence, decision making and so on. However, most research shows that practicing memorization of a single stimulus dimension with a single task leads to learning that is mostly specific to that task. Consistent with the learning principles discussed above, to whatever extent the color task may produce some benefits to working memory, greater benefits are likely to be achieved through training approaches that take into account the diverse uses of working memory and our knowledge of the neural underpinnings of working memory. For example, Deveau et al. (2015) proposed that training with a diversity of stimuli (e.g. instead of a single color, training with a diverse set of auditory and visual features) and task-structures (e.g. that address memory span, memory updating, memory resolution, etc) that are similar to the demands required of working memory in the world (e.g. understanding sentences, keeping track of daily tasks, problem solving, etc) is a needed step for achieving learning that transfers beyond the training materials.

Although it might seem ideal to train people on every possible use pattern they may encounter in the real world, this represents a very difficult feat to achieve. A new hypothesis being explored here is that there exists critical subset of cases, which we'll call a basis function, that are sufficiently characteristic of all the use patterns that may be encountered, such that training upon this basis is sufficient to achieve generalization to all others that the participant may experience in the world.

Seitz and colleagues (Deveau et al., 2015; Deveau, Lovcik, et al., 2014; Deveau, Ozer, et al., 2014; Deveau & Seitz, 2014) suggested the concept of a basis function as a core approach to brain training to achieve transfer to real world cognition. To specify a basis function it is necessary to identify a set of tasks and stimuli that spans the broad use of the cognitive skills in question. Understanding of the brain processes underlying those skills is an important constraint in choosing a candidate basis function. For example, in the case of vision, there is substantial evidence that the early visual system processes the visual world through a system of filters that approximate Gabor filters (Deveau & Seitz, 2014) of different spatial frequency, orientation and locations. Training using these stimuli has shown broad transfer to other domains including reading (DeLoss et al., 2015; Deveau, Lovcik, et al., 2014; Deveau, Ozer, et al., 2014; Deveau & Seitz, 2014; Polat, 2009).

The key question to ask in order to design training to meet a learning objective is: What are the set of challenges that can be presented to the learner that are appropriately representative of the target skill in the real world that, if trained, will be applied in the diversity of situations in the real world with greater success than if the training had not taken place. Then the training goal is to identify these challenges and that contain key “appropriately representative” characteristics and train the learner on these examples and then evaluate the extent to which this training generalizes to examples that are novel to the participants. While it is far from trivial to correctly identify what the true basis function for a given cognitive ability may be, the goal of any brain training approach is to arrive at the best possible approximation of such a basis function. This will inevitably be an iterative process where failures to find generalization will inform what cases may have been left out of the basis function and provide insight into what may be a new, expanded, training set.

While the idea of training a basis function may sound a little exotic, and identifying the appropriate dimensions of such a training is a non-trivial endeavor, this framework is similar to that required of any effective process of laying out learning objectives and designing training approaches that fully address those objectives. However, an advantage of the formalism of a basis function is that it provides an explicit approach of how to train and also the dimensions upon which adaptive evaluation during training should be targeted.

Continuous Evaluation During Training

An important feature of many brain training programs is their ability to continuously adapt task difficulty based upon the performance of the participant. For example, in the case of perceptual learning (Deveau, Ozer, et al., 2014), the visibility of task targets can be adjusted so that with correct responses the stimuli became more difficult to see and with incorrect responses the stimuli became easier to see. Similarly, in working memory training, the memory set-size (e.g., number of items that need to be remembered at once) is typically increased or decreased based upon high or low task accuracy, respectively. The goal of such adaptive procedures is to achieve a so-called Goldilocks effect (de Jonge et al., 2012) where difficulty during training is just right so as to maintain challenge and promote learning. While the exact attributes of this so-called Goldilocks point is more intuitive than quantifiable it relates to other research on flow (Csikszentmihalyi et al., 1992) where building evidence suggests that getting trainees in the right “zone” can lead to better training outcomes.

While some training programs may train a single skill, many brain training approaches involve training multidimensional constructs that represent different abilities or skills. In these cases adaptive procedures must estimate multiple participant abilities throughout the time course of training. For example, in the vision training described above, separate adaptive procedures estimated visual contrast sensitivity, visual acuity, ability to ignore visual distractors, and speed of processing which allowed each to be separately quantified throughout the course of training (Deveau, Lovcik, et al., 2014). The ability to simultaneously estimate the features of a high-dimensional learning space is a clear advantage of computerized training approaches.

Notably, the adaptive procedures used during training resemble, or in some cases are the same, as those used in tests of cognitive function to evaluate participants' abilities. However, a typical difference between training and testing is that training involves many thousands of trials sampled on multiple days spread across weeks, whereas testing may involve only a few hundred trials taken at just two time-points (before and after training). As such, training procedures actually provide a better sampling of proficiencies (at least on the training task) than is typically achieved through the outcome evaluation tests. Arguably, the use of many training sessions allows for more independent evaluations of skills than would be collected in typical testing procedures and as such provides for a more reliable data set.

This in fact resembles the situation described earlier of effective classroom teaching where there is a synthesis of testing and teaching. Effective lessons are best situated with knowledge of current student proficiencies and thus benefit from continuous evaluation by the teacher. This can be achieved through the adaptive procedures employed during training that moment by moment evaluate participant proficiencies in order to provide an appropriate next training challenge. We suggest that if such adaptive procedures were properly designed to validly assess that the learner has met learning objectives that they can serve to continuously estimate participant skills during training. For example, in the case of working memory training adaptive procedures are typically directed to increase or decrease the number of items that an individual is challenged to remember based upon the number of items that they successfully remembered in past trials. These data provide a very reliable measure of that participant's memory capacity.

We further suggest that if the continuous evaluation of participant skills during training increases the reliability of tests of participant proficiencies by combatting confounds of expectancy (often called placebo) effects. As stated earlier, a standard recommendation to improve the quality of brain training studies is to employ active control groups to measure expectation effects associated with participants knowing that they are in an experimental condition that is hypothesized to provide a benefit. However, expectancy effects manifest, at least in part, because participants know that the test sessions are important and so may put additional effort into them when they expect that they should do well. On the other hand, performance evaluated on a regular basis during training creates a circumstance where participants are less aware that they are being evaluated and task performance is likely to become more routine over time. Thus proficiencies estimated through repeated evaluation across training sessions may be less susceptible to expectancy effects. In other fields this method of "stealth testing" (J. V. Shute et al., 2016; V. Shute et al., 2017) is becoming increasingly popular where evaluation is embedded in games as a method evaluate participants while they are at ease.

We thus suggest that continuous evaluation of performance during training can provide a more reliable estimate of participant proficiencies than one-shot or isolated estimates found through conventional pre/post testing methodologies. Continuous evaluations can also minimize expectancy effects compared to standard pre/post testing approaches.

Use of valid measures of real world cognition

The next issues that we address with continuous evaluation of proficiencies during training are issues of test validity. Teaching to the test should be an even greater concern if the test is simply an estimate of proficiency during training. However, these issues of validity are very similar whether a separate testing session is used or whether testing is continuously conducted during training. The key, once again, is in explicitly defining the learning objectives, carefully designing tasks that measure performance in the activities critical to determining that those learning objectives are met (Mislevy, 2013) and repeatedly employing adaptive evaluation methods that ensure reliable measurement.

When designing a test battery to evaluate learning outcomes, the key considerations are: 1) creating a test battery that specifically measures each learning objective, and 2) using test items that differ from those used during training to evaluate transfer to new materials. These are also desirable attributes for continuous evaluation during training.

In regard to the first issue, this is where the concept of a basis function is most powerful. The demands of what should be contained in the test battery are essentially the same as that for the training set. Thus the adaptive algorithms used during training can be used to continuously sample performance across the tasks and stimuli employed during training in a manner that is more reliable than is easily achieved through pre/post testing.

Continuous Evaluation of Transfer

The second issue, of transfer to novel situations, is still of fundamental concern. Most testing regimes use different stimuli and tasks than used during training to validate whether benefits of training extends beyond the exact examples that were trained upon (Simons et al., 2016). While at first this may represent a conundrum to the approach of continuous evaluation, the solution is to vary the tasks and stimuli across the period of training. This both allows for a continuous evaluation of transfer during the course of training and also provides a more diverse training set that in turn is expected to produce greater transfer (Deveau & Seitz, 2014). For example, in the case of memory training, tests of memory capacity, ability to update memory, and the precision with which information is held in memory for a wide range of stimuli, of different modalities, over spatial and temporal dimensions, under load and distraction, and over different time-courses can be distributed throughout the training battery to classify changes of memory function (Deveau et al., 2015).

The key is that dimensions of transfer need to be temporally-distributed and tracked in a manner that clearly distinguishes at each point during the timecourse of training what measures are novel to the participant and which have been already presented during training. Each time a novel test item is encountered it serves as a measure of transfer. In addition to diversifying the training set throughout training, one can also directly insert test trials into the training to evaluate performance on other skills of interest, such as in stealth testing (J. V. Shute et al., 2016; V. Shute et al., 2017). For example the C8 training program, a new brain training approach that is used in schools to train attention, memory and decision making, inserts items from the NIH toolbox, a standardized set of cognitive testing tools, to collect

cognitive psychological tests interspersed with their main training program (Wexler et al., 2016).

The extent to which evaluation can be fully integrated into the training vs the extent to which some evaluation must necessarily be conducted in the more traditional pre/post tests may depend upon the case example. For example, when using a brain training tool to help a patient with ADHD, the most valid outcome measures will be how the patient performs at home, in school, at the workplace, etc., and these outcomes are not easily determinable within a computerized training tool (although see below). Still methods of continuous evaluation during training have the potential to achieve both more reliable and more valid evaluation than is typically achieved through computerized pre/post testing methodologies.

Can Adaptive Algorithms Trained on Transfer?

A novel idea afforded by this framework is that if one can continuously monitor transfer during training (as described above by measuring transfer whenever new example types are introduced) then this affords the possibility that one can potentially adapt training on not only performance on trained stimuli, but further on momentary transfer to untrained stimuli. For example, in cases where transfer is limited then it may make sense to introduce more variability in the training set. Likewise study of performance patterns during training of participants who show transfer of learning vs those who don't may yield additional strategies by which to adapt training to produce greater transfer. While this idea has not been explored to date, and research will be required to determine its potential, it is a natural extension proposed framework and may ultimately provide for approaches that will help identify when participants are adopting strategies that lead to specificity and move them towards strategies that will lead to greater transfer.

Testing in the Real world

A major limitation of the approaches discussed thus far is that most "laboratory" tests have questionable validity as being representative of real-world cognition. Essentially, through the course of computerized training, participants will become more familiar with the particular attributes of those tasks and there is a concern that improvements that extend broadly across variants of computerized testing will not extend to ecological contexts.

A potential solution to this problem is reality mining (Eagle et al., 2005), which utilizes mobile devices, social media, and other measured activities to understand participant behaviour. Nowadays, it is feasible to measure real world cognition through how people interact with these devices. For example, to test memory, one could measure how many times do people listen to a phone message to extract a phone number, how many times people check a grocery list when in a store, one can even examine GPS and step information of how many aisles are traversed at what rate in order to complete the shopping task, or ability to following driving directions. To test vision, what font size do they use on their phone, how often do they zoom into the text? With access of a camera, one can even look at head distance from screen to determine the participants reading acuity. Further, the microphone can be used to record conversations, actigraphy to record sleep and activity,

galvanic skin response and heart-rate recording stress/arousal while conducting different activities. While these methods sound exotic, and of course, privacy issues need to be addressed, this type of reality mining is an active area of research, and technology is increasingly embracing methods of addressing trade-offs of data-sharing and data privacy, especially in cases where users can gain a benefit from being monitored, such as in cases of health monitoring where benefits can be substantial.

Reality mining provides tremendous opportunity to measure participants' actual performance in activities of daily living and can be used both to evaluate skills during training as well as to measure training outcomes. However, many challenges exist in devising appropriate metrics, and evaluating their reliability and validity as predictors of participants' abilities in real world cognitive tasks. As such it may take some time before the potential gains of reality mining are fully realized.

Limitations of Continuous Evaluation

While we suggest that continuous evaluation during training can lead to better evaluation, and in turn better training outcome for brain training, the approach is not well-explored in the field and there are a number of limitations that need to be considered. A first issue, which has already been noted, is that transfer can only be measured within the limited domain of the training environment, as such many cases of transfer to real-world contexts will be impractical to measure in computer based training paradigms. Another issue is that continuous measurement of transfer during training does not allow for good comparison with active control conditions, which either need to introduce numerous testing trials that can themselves contribute to learning and in turn serve to diminish the magnitude of the observed difference between the intervention and control groups (Green, Strobach and Schubert, 2014), or would require use of a conventional pre/post test design for comparison. Likewise the use of continuous measure may not lend itself to good estimates of dose-responses (Jaeggi et al. 2008), again due to contamination by testing trials. As such while continuous evaluation affords a number of potential benefits, it will also provide challenges for traditional clinical trial structures by limiting good comparison to controls, estimates of dose-responses, and cannot address all components of transfer that may be relevant outcomes.

Conclusion

The use of continuous evaluation during training has substantial potential to achieve reliable and valid skill evaluation that addresses many of the limitations of using standard pre-test/post-test evaluation designs. Greater reliability is achieved through repeated testing over many days, allowing participants greater familiarity with testing procedures and at the same time allowing for establishment of stable baselines and evaluation of variability of testing proficiency over time. Validity can be improved by varying the test items over time in a manner that spans the relevant learning objectives. Expectancy effects are addressed by the removal of explicit test sessions. Further, the use of adaptive algorithms allows for personalization of evaluation that targets each individual with tests appropriate to their skill levels. In addition, employing mobile devices to track participant's normal behavior provides

an opportunity to capture additional information about proficiencies in real world cognitive tasks that are related to the learning objectives. While this synthesis of training and evaluation may be criticized as an extreme example of teaching to test, the main challenges, particularly in the choice of evaluation materials that are predictive of skill performance in real world activities, are largely the same as those of standard pre-test/post-test evaluation designs. Thus, arguably continuous evaluation during training provides advantages with few extra costs beyond the additional design requirements, over standard approaches in the field of brain training.

References

- Aberg KC, Tartaglia EM, Herzog MH. Perceptual learning with Chevrons requires a minimal number of trials, transfers to untrained directions, but does not require sleep. *Vision Research*. 2009; 49(16): 2087–2094. [PubMed: 19505495]
- Ahissar M, Hochstein S. Task difficulty and the specificity of perceptual learning. *Nature*. 1997; 387(6631):401–406. [PubMed: 9163425]
- Anguera JA, Boccanfuso J, Rintoul JL, Al-Hashimi O, Faraji F, Janowich J, Gazzaley A. Video game training enhances cognitive control in older adults. *Nature*. 2013; 501(7465):97–101. [PubMed: 24005416]
- Au J, Sheehan E, Tsai N, Duncan GJ, Buschkuehl M, Jaeggi SM. Improving fluid intelligence with training on working memory: a meta-analysis. *Psychon Bull Rev*. 2015; 22(2):366–377. [PubMed: 25102926]
- Baker CI, Peli E, Knouf N, Kanwisher NG. Reorganization of visual processing in macular degeneration. *J Neurosci*. 2005; 25(3):614–618. [PubMed: 15659597]
- Ball K, Sekuler R. Adaptive processing of visual motion. *J Exp Psychol Hum Percept Perform*. 1981; 7(4):780–794. [PubMed: 6457092]
- Beste C, Wascher E, Gunturkun O, Dinse HR. Improvement and impairment of visually guided behavior through LTP- and LTD-like exposure-based visual learning. *Curr Biol*. 2011; 21(10):876–882. [PubMed: 21549600]
- Boot WR, Blakely DP, Simons DJ. Do action video games improve perception and cognition? *Front Psychol*. 2011; 2:226. [PubMed: 21949513]
- Bul KC, Franken IH, Van der Oord S, Kato PM, Danckaerts M, Vreeke LJ, Maras A. Development and User Satisfaction of “Plan-It Commander,” a Serious Game for Children with ADHD. *Games Health J*. 2015; 4(6):502–512. [PubMed: 26325247]
- Bul KC, Kato KM, Van der Oord S, Danckaerts M, Vreeke LJ, Willems A, Maras A. Behavioral Outcome Effects of Serious Gaming as an Adjunct to Treatment for Children With Attention-Deficit/Hyperactivity Disorder: A Randomized Controlled Trial. *JOURNAL OF MEDICAL INTERNET RESEARCH*. 2016; 18(21):1–18.
- Chung ST. Improving reading speed for people with central vision loss through perceptual learning. *Invest Ophthalmol Vis Sci*. 2011; 52(2):1164–1170. [PubMed: 21087972]
- Crist RE, Kapadia MK, Westheimer G, Gilbert CD. Perceptual learning of spatial localization: specificity for orientation, position, and context. *J Neurophysiol*. 1997; 78(6):2889–2894. [PubMed: 9405509]
- Crist RE, Li W, Gilbert CD. Learning to see: experience and attention in primary visual cortex. *Nat Neurosci*. 2001; 4(5):519–525. [PubMed: 11319561]
- Csikszentmihalyi M, Rathunde K. The measurement of flow in everyday life: toward a theory of emergent motivation. *Nebr Symp Motiv*. 1992; 40:57–97. [PubMed: 1340523]
- Dahlin E, Nyberg L, Backman L, Neely AS. Plasticity of executive functioning in young and older adults: immediate training gains, transfer, and long-term maintenance. *Psychol Aging*. 2008; 23(4):720–730. [PubMed: 19140643]
- Das A, Tadin D, Huxlin KR. Beyond blindsight: properties of visual relearning in cortically blind fields. *J Neurosci*. 2014; 34(35):11652–11664. [PubMed: 25164661]

- de Jonge M, Tabbers HK, Pecher D, Zeelenberg R. The effect of study time distribution on learning and retention: a Goldilocks principle for presentation rate. *J Exp Psychol Learn Mem Cogn*. 2012; 38(2):405–412. [PubMed: 22004269]
- DeLoss DJ, Watanabe T, Andersen GJ. Improving vision among older adults: behavioral training to improve sight. *Psychol Sci*. 2015; 26(4):456–466. [PubMed: 25749697]
- Deveau J, Jaeggi S, Zordan V, Phung C, Seitz AR. How to build better memory training games. *Frontiers in Systems Neuroscience*. 2015; 8(243):1–7.
- Deveau J, Lovcik G, Seitz AR. Broad-based visual benefits from training with an integrated perceptual-learning video game. *Vision Res*. 2014
- Deveau J, Ozer DJ, Seitz AR. Improved vision and on-field performance in baseball through perceptual learning. *Curr Biol*. 2014; 24(4):R146–147. [PubMed: 24556432]
- Deveau J, Seitz AR. Applying perceptual learning to achieve practical changes in vision. *Front Psychol*. 2014; 5:1166. [PubMed: 25360128]
- Diziol D, Walker E, Rummel N, Koedinger KR. Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review*. 2010; 22(1):89–102.
- Dovis S, Van der Oord S, Wiers RW, Prins PJ. Improving executive functioning in children with ADHD: training multiple executive functions within the context of a computer game. a randomized double-blind placebo controlled trial. *PLoS ONE*. 2015; 10(4):e0121651. [PubMed: 25844638]
- Drew T, Vo ML, Wolfe JM. The invisible gorilla strikes again: sustained inattention blindness in expert observers. *Psychol Sci*. 2013; 24(9):1848–1853. [PubMed: 23863753]
- Eagle N, Pendland AS. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*. 2005; 10(4):255–268.
- Fahle M. Perceptual learning: specificity versus generalization. *Curr Opin Neurobiol*. 2005; 15(2):154–160. [PubMed: 15831396]
- Fahle, M., Poggio, T. Perceptual learning. Cambridge, Mass: MIT Press; 2002.
- Fiorentini A, Berardi N. Perceptual learning specific for orientation and spatial frequency. *Nature*. 1980; 287(5777):43–44. [PubMed: 7412873]
- Franceschini S, Gori S, Ruffino M, Viola S, Molteni M, Facoetti A. Action video games make dyslexic children read better. *Curr Biol*. 2013; 23(6):462–466. [PubMed: 23453956]
- Green CS, Bavelier D. Action video game modifies visual selective attention. *Nature*. 2003; 423(6939):534–537. [PubMed: 12774121]
- Green CS, Bavelier D. Action-video-game experience alters the spatial resolution of vision. *Psychol Sci*. 2007; 18(1):88–94. [PubMed: 17362383]
- Green CS, Strobach T, Schubert T. On methodological standards in training and transfer experiments. *Psychological Research*. 2014; 78(6):756–772. [PubMed: 24346424]
- Hardy JL, Nelson RA, Thomason ME, Sternberg DA, Katovich K, Farzin F, Scanlon M. Enhancing Cognitive Abilities with Comprehensive Training: A Large, Online, Randomized, Active-Controlled Trial. *PLoS ONE*. 2015; 10(9):e0134467. [PubMed: 26333022]
- Hikosaka O, Nakamura K, Sakai K, Nakahara H. Central mechanisms of motor skill learning. *Curr Opin Neurobiol*. 2002; 12(2):217–222. [PubMed: 12015240]
- Hofmann W, Schmeichel BJ, Baddeley AD. Executive functions and self-regulation. *Trends Cogn Sci*. 2012; 16(3):174–180. [PubMed: 22336729]
- Huang CB, Zhou Y, Lu ZL. Broad bandwidth of perceptual learning in the visual system of adults with anisometric amblyopia. *Proc Natl Acad Sci U S A*. 2008; 105(10):4068–4073. [PubMed: 18316716]
- Hung SC, Seitz AR. Prolonged training at threshold promotes robust retinotopic specificity in perceptual learning. *J Neurosci*. 2014; 34(25):8423–8431. [PubMed: 24948798]
- Huxlin KR, Martin T, Kelly K, Riley M, Friedman DI, Burgin WS, Hayhoe M. Perceptual relearning of complex visual motion after V1 damage in humans. *J Neurosci*. 2009; 29(13):3981–3991. [PubMed: 19339594]
- Jaeggi SM, Buschkuhl M, Jonides J, Perrig WJ. Improving fluid intelligence with training on working memory. *Proc Natl Acad Sci U S A*. 2008; 105(19):6829–6833. [PubMed: 18443283]

- Jeter PE, Doshier BA, Liu SH, Lu ZL. Specificity of perceptual learning increases with increased training. *Vision Res.* 2010; 50(19):1928–1940. [PubMed: 20624413]
- Jeter PE, Doshier BA, Petrov A, Lu ZL. Task precision at transfer determines specificity of perceptual learning. *J Vis.* 2009; 9(3):11–13.
- Klingberg T, Fernell E, Olesen PJ, Johnson M, Gustafsson P, Dahlstrom K, Westerberg H. Computerized training of working memory in children with ADHD—a randomized, controlled trial. *J Am Acad Child Adolesc Psychiatry.* 2005; 44(2):177–186. [PubMed: 15689731]
- Klingberg T, Forssberg H, Westerberg H. Training of working memory in children with ADHD. *J Clin Exp Neuropsychol.* 2002; 24(6):781–791. [PubMed: 12424652]
- Li J, Thompson B, Deng D, Chan LY, Yu M, Hess RF. Dichoptic training enables the adult amblyopic brain to learn. *Curr Biol.* 2013; 23(8):R308–309. [PubMed: 23618662]
- Li RW, Ngo C, Nguyen J, Levi DM. Video-game play induces plasticity in the visual system of adults with amblyopia. *PLoS Biol.* 2011; 9(8):e1001135. [PubMed: 21912514]
- Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P. Temporal processing deficits of language-learning impaired children ameliorated by training. *Science.* 1996; 271(5245):77–81. [PubMed: 8539603]
- Mishra J, de Villiers-Sidani E, Merzenich M, Gazzaley A. Adaptive training diminishes distractibility in aging across species. *Neuron.* 2014; 84(5):1091–1103. [PubMed: 25467987]
- Mislevy RJ. Evidence-centered design for simulation-based assessment. *Mil Med.* 2013; 178(10 Suppl):107–114.
- O'Rourke, E., Andersen, E., Gulwani, S., Popovi, Z. A Framework for Automatically Generating Interactive Instructional Scaffolding. Paper presented at the 33rd Annual ACM Conference on Human Factors in Computing Systems; Seoul, Republic of Korea. 2015.
- Owen AM, Hampshire A, Grahn JA, Stenton R, Dajani S, Burns AS, Ballard CG. Putting brain training to the test. *Nature.* 2010; 465(7299):775–778. [PubMed: 20407435]
- Piaget, J. The origins of intelligence in children. New York: International Universities Press; 1952.
- Polat U. Making perceptual learning practical to improve visual functions. *Vision Res.* 2009; 49(21):2566–2573. [PubMed: 19520103]
- Polat U, Ma-Naim T, Belkin M, Sagi D. Improving vision in adult amblyopia by perceptual learning. *Proc Natl Acad Sci U S A.* 2004; 101(17):6692–6697. [PubMed: 15096608]
- Polat U, Schor C, Tong JL, Zomet A, Lev M, Yehezkel O, Levi DM. Training the brain to overcome the effect of aging on the human eye. *Sci Rep.* 2012; 2:278. [PubMed: 22363834]
- Popov T, Jordanov T, Rockstroh B, Elbert T, Merzenich MM, Miller GA. Specific cognitive training normalizes auditory sensory gating in schizophrenia: a randomized trial. *Biol Psychiatry.* 2011; 69(5):465–471. [PubMed: 21092939]
- Raiguel S, Vogels R, Mysore SG, Orban GA. Learning to see the difference specifically alters the most informative V4 neurons. *J Neurosci.* 2006; 26(24):6589–6602. [PubMed: 16775147]
- Sagi D, Tanne D. Perceptual learning: learning to see. *Curr Opin Neurobiol.* 1994; 4(2):195–199. [PubMed: 8038576]
- Schoups A, Vogels R, Qian N, Orban G. Practising orientation identification improves orientation coding in V1 neurons. *Nature.* 2001; 412(6846):549–553. [PubMed: 11484056]
- Schoups AA, Vogels R, Orban GA. Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularly. *J Physiol.* 1995; 483(Pt 3):797–810. [PubMed: 7776259]
- Seitz A. Primer: Perceptual Learning. *Current Biology.* 2017 in press.
- Seitz AR, Kim R, Shams L. Sound facilitates visual learning. *Curr Biol.* 2006; 16(14):1422–1427. [PubMed: 16860741]
- Seitz AR, Watanabe T. The phenomenon of task-irrelevant perceptual learning. *Vision Res.* 2009; 49(21):2604–2610. [PubMed: 19665471]
- Shams L, Seitz AR. Benefits of multisensory learning. *Trends Cogn Sci.* 2008
- Shipstead Z, Redick TS, Engle RW. Is working memory training effective? *Psychol Bull.* 2012; 138(4):628–654. [PubMed: 22409508]

- Shute JV, Leighton JP, Jang EE, Chu MW. Advances in the Science of Assessment. Educational Assessment. 2016; 21(1):34–59.
- Shute, V., Ke, F., Wang, L. Assessment and Adaptation in Games. In: Wouters, P., Oostendrop, HV., editors. Instructional Techniques to Facilitate Learning and Motivation of Serious Games. Switzerland: Springer; International Publishing; 2017. p. 59-78.
- Simons DJ, Boot WR, Charness N, Gathercole SE, Chabris CF, Hambrick DZ, Stine-Morrow EA. Do “Brain-Training” Programs Work? Psychol Sci Public Interest. 2016; 17(3):103–186. [PubMed: 27697851]
- Smith GE, Housen P, Yaffe K, Ruff R, Kennison RF, Mahncke HW, Zelinski EM. A cognitive training program based on principles of brain plasticity: results from the Improvement in Memory with Plasticity-based Adaptive Cognitive Training (IMPACT) study. J Am Geriatr Soc. 2009; 57(4): 594–603. [PubMed: 19220558]
- Szpiro SF, Wright BA, Carrasco M. Learning one task by interleaving practice with another task. Vision Res. 2014; 101:118–124. [PubMed: 24959653]
- Thorndike, EL. Educational Psychology. New York: Teachers college, Columbia University; 1913.
- Titz C, Karbach J. Working memory and executive functions: effects of training on academic achievement. Psychol Res. 2014; 78(6):852–868. [PubMed: 24389706]
- Vlahou EL, Protopapas A, Seitz AR. Implicit training of nonnative speech stimuli. J Exp Psychol Gen. 2012; 141(2):363–381. [PubMed: 21910556]
- Volante L. Teaching To the Test: What Every Educator and Policy-maker Should Know. Canadian Journal of Educational Administration and Policy(n35). 2004
- Watanabe T, Nanez JE Sr, Koyama S, Mukai I, Liederman J, Sasaki Y. Greater plasticity in lower-level than higher-level visual motion processing in a passive perceptual learning task. Nat Neurosci. 2002; 5(10):1003–1009. [PubMed: 12219093]
- Wexler BE, Iseli M, Leon S, Zaggle W, Rush C, Goodman A, Bo E. Cognitive Priming and Cognitive Training: Immediate and Far Transfer to Academic Skills in Children. Sci Rep. 2016; 6:32859. [PubMed: 27615029]
- Whitton JP, Hancock KE, Polley DB. Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise. Proc Natl Acad Sci U S A. 2014; 111(25):E2606–2615. [PubMed: 24927596]
- Xiao LQ, Zhang JY, Wang R, Klein SA, Levi DM, Yu C. Complete Transfer of Perceptual Learning across Retinal Locations Enabled by Double Training. Curr Biol. 2008; 18(24):1922–1926. [PubMed: 19062277]
- Zhou J, Zhang Y, Dai Y, Zhao H, Liu R, Hou F, Zhou Y. The eye limits the brain’s learning potential. Sci Rep. 2012; 2:364. [PubMed: 22509464]