The Role of the Continuous Performance Test

by Siegfried Othmer, Ph.D.

We have in neurofeedback a physiologically-based tool which should, in the best of circumstances, be accompanied by physiologically-based assessments in addition to the usual symptom-tracking. A case can be made for tracking measures that monitor the training process as well as for pre-post assessments. These assessments can either consist of passive monitoring of physiological variables or of active functional challenges. Typically both of these approaches are handicapped by intrinsic variability in the measures. When it comes to real-time tracking measures, our response to lack of reliability is to employ several independent measures, and to look for consistency among them. When it comes to pre-post measurements, we look to repeated measures to gain statistical precision.

The continuous performance test has been devised to yield a reliable pre-post measure on attentional variables. It is a go/no-go challenge test, or what is known as a choice reaction time test. In the case of the QIKtest, which emulates the TOVA (the Test of Variables of Attention ®), the target and non-target discrimination only requires a knowledge of up versus down.

The test challenges the person under both high-demand and low-demand conditions. Under high-load conditions, the testee is more likely to make errors of commission, whereas under low-load conditions, the testee is more likely to make errors of omission. In addition to counting errors of omission and commission, the test determines the average reaction time, as well as the standard deviation of reaction time, the variability. The test duration is some 21 minutes, and in the case of the QIKtest involves five sequential trials of under-load and over-load conditions, staged in the sequence as follows: low, low, high, high, low. Each segment lasts for about four minutes.

The functional adult brain is highly unlikely to make errors of omission, so a lengthy test has to be allowed for in order to get adequate statistics on omission errors. Identical conditions are maintained over each four-minute segment so that nothing could serve as an additional alerting stimulus. The ratio of go to no-go stimuli is 7:2 in the high-demand segments, and the reverse in the low-demand segments. With few responses called for in the low-demand segments, this part of the test challenges the person's ability to maintain vigilance even in the face of a boring and invariant task. The low-demand phase is divided into two segments in order to examine whether the person is able to accommodate to this challenge or whether performance degrades over time.

The high-demand phase of the test is similarly divided into two segments to test whether performance improves or degrades over time. When the ratio of go to no-go cues suddenly goes from 2:7 to 7:2, many subjects initially do poorly, which may be ascribed to an anxiety or even freezing response. Subsequently they may recover and end up doing well. Others will see their performance degrade over time as boredom once again takes over.

The QIKtest distinguishes itself from the TOVA in that it has added a fifth segment to investigate the transition from high-load back to low-load conditions, as well as to allow comparison under the same conditions of the beginning and the end of the test.

Continuous performance tests have shown excellent test-retest reliability, and are thus very good candidates as change measures in neurofeedback. At the same time, they are also sensitive tests, so a child's test results will depend on fatigue level, and hence time of day. The test may reflect a person's poor night of sleep and perhaps even whether a child had breakfast that morning. In order to assure that the test results reflect the neurofeedback training, the therapist should arrange for all the tests to be taken under comparable conditions. The norms are intended for testing during the morning hours, i.e. before lunch, and performance degradation is to be expected if children are tested after school.

Given the emphasis in our assessments on a person's functioning in the arousal domain, and along the arousal axis, the fact that the CPT evaluates both high- and low-load conditions makes it particularly suitable for our work. The test can be very revealing with regard to what kind of depression the person may be suffering, and where the specific deficits lie in a child who may be referred for ADHD. Our collective experience is that recovery of good function on the CPT is within the capability of most nervous systems, even of those that may have suffered a traumatic brain injury. Hence we have very high expectations for normalizing the CPT response in nearly all of our clients.

When the CPT performance is not yet normalized, it is usually an indication that more work needs to be done. Conversely, normalization of the QIKtest scores does not indicate that the potential of neurofeedback has been exhausted for a particular person. Finally, a person coming in with normal CPT scores may still benefit from neurofeedback. The data indicate that with neurofeedback a person may end up scoring well above prevailing norms in terms of impulsivity, reaction time, and variability. Matters are different in the case of omission errors, since the expectation is that a mature nervous system will not make such errors. This leaves no headroom for scores to improve.

Conventionally only the statistical data derived from the CPT are used in assessments, since these can readily be compared to norms. The CPT can yield additional useful data, however. In the QIKtest report we include the results of all individual trials throughout all the segments. The 21-minute test, with an inter-stimulus interval of two seconds, can be thought of as a sequence of 630 individual challenges, half of which call for a motor response. Not only are overall trends of interest, but individual results as well. First of all, as Larry Greenberg (designer of the TOVA) recognized, people react in a characteristic fashion after making a commission error. Either they rouse themselves to an even greater state of arousal and react even faster on the next trial, or they realize that they must act with more deliberation and slow down their next response.

It is also of interest to look for outliers among the responses. These may be few enough in number that they don't significantly degrade the overall score, and yet they may indicate episodes where the nervous system was not responsive. In this case, the outliers may make the case for additional training even though the overall scores may have normalized. The outliers may also be the result of the nervous system undergoing a paroxysmal event. If such a paroxysmal event is sufficiently brief, and does not result in loss of consciousness, its existence may very well be missed in other

assessments. If these events are more frequent, but yet clearly distinguishable from well-behaved performance otherwise, the paroxysmal activity may well degrade function such as sensory processing. The CPT may be the first evidence that such a problem exists.

Advantages of the QIKtest

The QIKtest has implemented the same testing philosophy that guided the TOVA. Given the long and successful history of the test, it should not be altered without good cause. Also, there is value in maintaining continuity of a test over time for comparability across generations. At the hardware level, the use of a microprocessor-based system allows absolute timing integrity at the level of 0.1 milliseconds. Secondly, programmability allows the QIK CPT to be complemented by other such challenges in the future.

Other hardware features include an auditory output capability, so that the same test can be given with an auditory instead of a visual stimulus. They also include an output for timing signals to be provided to an EEG-monitoring instrument. In this manner, performance can be correlated with EEG measures, and evoked potential measurements are facilitated.

At the level of the analysis software, we altered the assumption that any reaction time of less than 200 ms had to be regarded as an anticipatory response. There were too many cases over time where these presumptive anticipatory responses were quite in character for the person, showing up as the tail of Gaussian distribution of responses, and where these fast responses were found to be systematically correct. They could not have been random hits. A new criterion of 160ms was installed in the QIK in place of the 200msec of the TOVA.

The QIKtest report is web-based, and it features full graphical description of the entire response history for each test. A variety of statistical analyses are presented first, followed by the graphical representations, and finally a summary of tests for the individual to facilitate comparison across the training history. The report can be printed out in black-and-white or as a PDF file in color.

The versatility of the web-based scoring allows us certain options in terms of norming. For example, the original norms can be updated to the present day. Current norms are expected to be different because children are almost uniformly now exposed to video games and to faster image transitions on TV shows. It is also possible to construct supernorms that reflect good function more than merely the ambient population-based norms. Further, it is possible to move to non-Gaussian statistics, since the distributions for the four sub-categories deviate so significantly from a Gaussian distribution.

Marketing benefit of the CPT

Whenever greater care is taken in an assessment, the family also takes the results more seriously. Often the CPT data helps to persuade the family that there is a "real" issue to be pursued with neurofeedback (as opposed to mere oppositionality in the child, for example, that may be deemed a purely psychological issue). Similarly, when progress is shown on the CPT at retest, the parents are more likely to credit the neurofeedback for the other improvements they have observed in their child. The concreteness of the computer-scored CPT data will have made the case.

Controversies Concerning the CPT

The controversies that surround the continuous performance test relate essentially to the various interpretations of the data to serve different purposes. The questions that relate to the test itself have been answered affirmatively. That is to say, the test gives reliable, stable and repeatable results for a nervous system that is itself stable. Sufficient samples are provided for to yield statistically robust findings (with the singular exception already mentioned: omission errors in functional adults, where errors are rare). And there is no observable practice effect. Once the first-level interpretive data are in hand, however, experts go their different ways in interpreting the data.

The developers of the TOVA use it mainly for titration of stimulant medication. They have observed, for example, that the performance peak achievable with stimulants may correspond to a fairly narrow range in medication dosage. Moreover, the performance peak from the standpoint of attentional functioning may differ from that related to behavioral control. The latter is typically higher. The TOVA derives an overall ADHD score from the data, which can be tracked over a range of medication.

Clearly the ADHD score is only intended to complement other aspects of the overall assessment, but once such a score is calculated the discussion of false positives and false negatives becomes unavoidable. This is a shame because that whole discussion, which can never be finally put to rest, then reflects negatively back onto the basic test. To us, the issue is straight-forward. The CPT measures certain attentional variables under controlled conditions. There are no distractors, for example. So there is no question here of replicating actual life situations in which ADHD children often manifest their problems. No single test can meet that burden. The point is that if the CPT scores indicate a deficit even under these relatively ideal circumstances, then a functional neurophysiological deficit is likely indicated. (And our ability to normalize such function purely with a training paradigm supports that understanding.)

The Conners CPT and the IVA CPT both interpret the basic data in terms of a large number of behavioral categories. The specificity and reliability of these subsidiary classifications have never been established in either case. This just enlarges the playing field on which clinicians and researchers may then disagree. Most clinicians probably just think of these subclassifications as having some heuristic value, nothing more, which keeps these disagreements from becoming heated.

One approach to these controversies is to take the same attitude as has been recommended with regard to IQ tests: It is simply asserted that for purposes of allowing the conversation to progress, "IQ is what the IQ test tests." This takes the discussion closer to the data at hand, and tries to finesse the basic controversy. Our inclination is to take this one step further. We are actually interested in precisely what is measured here as a probe of nervous system function—of its ability to maintain vigilance under challenging conditions; the existence of outlier responses; the consistency of performance across short and long time scales, etc. We are interested in all the individual pieces of data. There is no need to reintegrate them all into one conceptual entity.

We don't need to bring up ADHD at all. The diagnostic threshold is not relevant for purposes of neurofeedback in any event. And in ADHD we have a concept that is far more murky, diffuse and multi-factorial than what we started out with in the CPT. So the usual argument of indicting the quantitative CPT with the far less quantifiable entity of ADHD has it entirely backwards. The CPT

represents solid behavioral data that needs to be accommodated in any other model; it cannot be dismissed by reference to something that is amorphous, ambiguous, and bereft of quantitative handholds and definable boundaries.

This approach finds support in the empirical finding that clinical populations do not distinguish themselves readily from the ADHD population when it comes to the CPT. We perform CPT tests on all clients who are capable of handling the test, and if one looks at all the CPT data collectively for anxiety, depression, headache, sleep disorders, and the ADHD spectrum it is not possible to tell one data set from another. Attentional failure is common to them all, and the underlying failure mechanisms are probably the same in all.

Generalizability of the CPT data

The broad utility of the CPT as a progress measure in neurofeedback relates to the fact that in this test we are laying bare some essential qualities of the functioning of the nervous system. The attentional set required to perform this task can be thought of as a particular state of the neural networks. One question then relates to the ability to maintain this attentional set as a subroutine while the brain otherwise entertains itself during the extended test interval; i.e., in the presence of internal distractors. A second question relates to the microscopic consistency with which a response is executed when called for. And a third question relates to the ability to inhibit a motor response under the duress of fast responding.

The entire cerebral architecture is organized around the chain of events from sensory inputs to motor output, all under the management of executive function. Even the execution of a simple motor response, then, samples the integrity of our nervous system quite broadly. And the determinants of good function in the sampled networks are not unique to these networks, but rather are more universal aspects of neural network organization. From the perspective of neurofeedback, these determinants of good function relate to the quality of communication within and between neural systems, which at the operational level is predominantly a matter of timing. The neurofeedback challenge affects the organization of timing relationships directly, and it does so in a fairly general and non-specific way. The motor act simply affords us a convenient behavioral observable with which we can quantify the consequences.

Hence it is no surprise that as performance on the CPT is observed to improve in neurofeedback there should be other improvements as well. Not only should we reject the intimate connection between CPT tests and ADHD that is our historical legacy. We should even look beyond the immediate issues of sensory processing, of executive function, and of motor control that are under test. Not only does neurofeedback affect neural network function more broadly, but also the CPT has broader implications than for the functions that it explicitly tests.

Complementary Tests with the QIK

Since the CPT takes 21 minutes to administer and is quite a chore to take, this is not something one wants to inflict on a client too often. A quicker test is needed for an ongoing progress measure. We have devised a one-minute test in which the challenge is progressive throughout. So it is more

likely to test a person's limits. The test is adequate to get a sense of the person's reaction time and variability. We may even get some omission and commission errors, but we don't expect these to relate to the corresponding measures in the CPT. The test is not boring, first of all, and these errors would be generated under very different conditions than in the CPT. Also, the statistics on these will be poor because of the brevity of the test. Nevertheless, a learning curve should be demonstrable if the test is done at every session.

It is recommended that only one such test be given at every session, and that it be given before the neurofeedback training. The test should not be used both before and after a session to demonstrate progress because such progress may well be obscured by the fatigue factor attributable to the training itself. It is more likely than not that clients will test worse after a session, and this cannot be taken as an indictment of either the training generally or the protocol specifically.

Quantitative EEG Measures with the QIK

The EEG is potentially much more revealing when investigated under challenge than it is under baseline conditions. It is of interest to look at the EEG on three different time scales in this regard. The first simply involves registering the usual band amplitudes over baseline and test conditions. Tracking the EEG under the challenge of a CPT can be thought of as an analogue of the physiological stress profiling that is commonly done in peripheral biofeedback, or it could simply be a constituent of such a profile that also looks at peripheral physiology at the same time.

The EEG under challenge falls into two broad categories for the compromised brain. The EEG may tend to normalize under the challenge, summoning its resources, or it may succumb to stress, fatigue, or somnolence. In either case the failure modes of the brain will be discernible.

A second time scale of interest is the four-second interval in which a particular trial is imbedded. If outlier responses are observed, it may be possible to discern the associated paroxysmal activity. Unfortunately, if none is observed it cannot be concluded that none exists. But if such events are detected then they can be targeted in an inhibit strategy.

The third timescale is that of the evoked potential, where we are looking at time resolution of milliseconds in a total interval of about one second post-stimulus. Since the event-related potential, or ERP, is dependent on the ambient EEG spectral components at the time of stimulus presentation (e.g., the phase of the ongoing alpha activity), one must also capture pre-stimulus data. For the latter two applications, the timing reference from the QIK is needed.

Summary

The CPT test may well be the most revealing and efficient test available to accompany neurofeedback as a progress measure. Among the CPTs available, the design philosophy implicit in the TOVA is the most suitable for our purposes. The QIK updates the TOVA to modern technical requirements and allows open-ended future developments. Finally, it allows complementary EEG measurements to be made that allow more subtle failure modes to be identified.