Summary of IVA-QS Research Studies

The IVA+Plus and IVA-QS test procedures, structure and quotient scales are essentially the same. Howeverr, the IVA-QS differs in that it is limited to the first 300 trials and, consequently, has a different set of norms. The IVA-QS is shorter and takes only about 8 minutes. It was designed specifically as a screening instrument. Research studies pertaining to its validity and reliability are discussed below.

IVA-QS Test Construct Validity

The IVA-QS has a construct validity based on the well establish Continuous Performance Test (CPT) model. This model has been extensively researched and when fully implemented consists of both high and a low demand components. In high demand conditions the test taker is required to rapidly respond to targets that are frequently presented with occasional non-targets. The primary purpose of the high demand test phase is to measure impulsivity (i.e., clicking to the non-target proceeded by a series of targets). In contrast, the low demand condition is incorporated in order to assess inattention. In the low demand condition the non-targets are frequent and targets rarely presented; requiring the test taker to sustain attention throughout this inherently boring task. Both high and low demand conditions are included in the IVA-QS test construction and they alternate every 50 trials. In addition, the IVA-QS like the IVA-2 integrates both visual and auditory test stimuli and, thus, does not require the administration of two separate sensory modality tests. Whereas, many individuals taking a CPT can learn to expect after a number of targets of the same sensory modality are presented that soon there will be a non-target and vice versa for the low demand condition, it is extremely challenging to apply this type of test taking strategy to the IVA-QS when presented with inter-mixed visual and auditory test stimuli. This unique construct design of the IVA-QS forces each person to pay close attention to each test item and any wavering of attention is very likely to result in either an error of impulsivity or inattention. This unique test construction design contributes to the high level of sensitivity and specificity found in the research studies discussed below.

Concurrent Validity

The IVA-QS and IVA+Plus scores and resulting diagnoses were evaluated for concurrent validity with one another. As noted above, the primary difference between the two tests is that the IVA-QS is shorter than the IVA+Plus. It consists of 300 trials versus the 500 trials that make up the latter. Two analyses were completed to examine the concurrent validity between the IVA+Plus and the IVA-QS. These were completed in order to determine if the IVA-QS (Quick Screen version) test scores were comparable to the matching test scores for the IVA+Plus. In the first analysis, the correlations between the primary and global scale test scores of each test were compared using the normative database. A second analysis was completed to determine if the IVA-QS resulted in the same diagnostic classification for individuals with ADHD diagnoses in comparison to the IVA+Plus using the appropriate diagnostic flowchart for each test.

The normative database used in this study consists of 1700 individuals (903 males and 797 females) ranging in age from six to 96 years old. The normative database records were de-identified. This normative database is used to calculate the norms for both the IVA+Plus and the IVA-QS. The difference between these two sets of norms is that the IVA-QS norms are based only on the first 300 trials of the test and also exclude the cool-down separate test component of the IVA+Plus.

Normative data was collected from individuals not known to have past neurological disorders, not be taking any medication aside from birth control and nasal sprays, not currently active in psychotherapy or counseling, not known to have learning or attentional problems, and not known to demonstrate hyperactivity. The purpose of using a reference group without any known influences that would impair test performance was to help make the IVA-QS CPT more sensitive to impairments in mental processing problems. This normative database was collected from a number of different sites throughout the country.

All of the Pearson correlation coefficients for the primary quotient scale scores were found to be significant and positively correlated. The correlations for these scales are presented below in Table 1-1 titled Correlations between IVA+Plus and IVA-QS Primary Scale Quotient Scores. All IVA-QS global quotient scale scores were also found to be significantly and positively correlated. The results of this analysis are presented in Table 1-2 below.

Almost all of the correlations were found to have very strong effect sizes. These results support that the IVA-QS demonstrates very strong concurrent validity with the established IVA+Plus test. The only noteworthy difference between the IVA-QS and the IVA+Plus was that the Stamina scale was relatively less in terms of its effect size. This difference is not considered unusual given that the IVA+Plus test is about 5 minutes

longer. Its impact is minimal, because the effect size is significant and reflects moderate strength. Consequently, the IVA-QS can be used with confidence in assessing and measuring attention and response control.

Thirty clients previously diagnosed by Dr. Sandford as having a primary diagnosis of ADHD were selected for the second concurrent validity evaluation. These clients had volunteered and given permission to have their clinical test data used in this study. These data records were de-identified. The patient data used in this study were selected from patients seen between the years of 2001 and 2011. Fifteen of these individuals were male, and fifteen were female. The ages ranged from 6 to 55, and the mean age was 17.1. Thirty individuals from the normative database, matched by age and sex, were randomly selected for comparison with this group. None of these individuals from the normative group were identified as having been diagnosed with ADHD, having ADHD-type symptoms, or having any other factors likely to impair their test functioning. The ADHD diagnoses derived from the IVA-QS ADHD were compared to the ADHD diagnoses derived from the IVA+Plus.

This comparison is useful in establishing the concurrent validity of the IVA-QS. The diagnosis derived from the IVA-QS test results agreed with the IVA+Plus for 85% of the individuals tested. When the IVA+Plus diagnosis was ADHD, the IVA-QS diagnosis supported a diagnosis of ADHD for 83% of the cases. In cases when the IVA+Plus diagnosis did not identify ADHD, the IVA-QS diagnosis was in agreement for 87% of the cases. These findings establish the concurrent validity of the IVA-QS test and are presented in Table 1-3a and Table 1-3b below. Further analysis of the accuracy of the IVA-QS in diagnosing ADHD separately and in combination with ADHD rating scales is provided in a diagnostic validity study below.

Table 1-1. Correlations between IVA+Plus and IVA-QS Primary Scale Quotient Scores

Primary Scale	r	p-value	Effect Size
Auditory Prudence (PRA)	0.91	< 0.001	Very Strong
Visual Prudence (PRV)	0.92	< 0.001	Very Strong
Auditory Consistency (CONA)	0.90	< 0.001	Very Strong
Visual Consistency (CONV)	0.89	< 0.001	Very Strong
Auditory Stamina (STMA)	0.55	< 0.01	Moderate
Visual Stamina (STMV)	0.59	< 0.01	Moderate
Auditory Vigilance (VIA)	0.77	< 0.001	Strong
Visual Vigilance (VIV)	0.81	< 0.001	Very Strong
Auditory Focus (FOCA)	0.92	< 0.001	Very Strong
Visual Focus (FOCV)	0.89	< 0.001	Very Strong
Auditory Speed (MNA)	0.97	< 0.001	Very Strong
Visual Speed (MNV)	0.98	< 0.001	Very Strong

Table 1-2. Correlations between IVA+Plus and IVA-QS Global Scale Quotient Scores

Global Scale	r	p-value	Effect Size
Full-Scale Response Control (FRCQ)	0.83	< 0.001	Very Strong
Auditory Response Control (ARCQ)	0.80	< 0.001	Very Strong
Visual Response Control (VRCQ)	0.81	< 0.001	Very Strong
Full-Scale Attention (FAQ)	0.89	< 0.001	Very Strong
Auditory Attention (AAQ)	0.87	< 0.001	Very Strong
Visual Attention (VAQ)	0.87	< 0.001	Very Strong
Sustained Auditory Attention (SAAQ)	0.85	< 0.001	Very Strong
Sustained Visual Attention (SVAQ)	0.88	< 0.001	Very Strong

Table 1-3a. Diagnostic Comparison of IVA-QS and IVA+Plus for ADHD Diagnoses

		IVA+Plus Test Diagnoses		
		ADHD	No ADHD	Total
IVA-QS Test Diagnosis	ADHD	24	4	28
	No ADHD	5	27	32
	Total	29	31	60

Table 1-3b. Diagnostic Comparison of IVA-QS and IVA+Plus for ADHD Diagnoses

	Diagnostic Accuracy
Sensitivity	83%
Specificity	87%
Positive Predictive Power	86%
Negative Predictive Power	84%

Diagnostic Validity Study

The clinical usefulness and diagnostic validity of computerized visual continuous performance tests (CPT) in the assessment and diagnosis of Attention Deficit-Hyperactivity Disorder (ADHD) has been called into question in presentations by two leading researchers (Goldman, 1994 and Halperin, 1994). Both of these researchers had concluded that the diagnostic validity of current visual computerized CPTs was not sensitive or specific to the degree that this type of test should normally be included in a multi-method assessment battery for ADHD. Barkley (1990) advanced his view that the potential for the computerized CPT was great, because this part of a comprehensive assessment was not tainted by the personal opinion biases that can occur in subjective rating scales. The problem in using a CPT arises at accurately interpreting its findings at the individual level. Barkley (1994a) in a reanalysis of earlier published data reported that a visual CPT correctly classified over 90% of the children with an abnormal score, but had false negative rates of 37% or higher. In other words, a visual CPT failed to identify about 2 of 5 children previously diagnosed using comprehensive assessment techniques for diagnosing ADHD.

To date, most computerized CPTs are visual, leaving out the assessment of possible auditory impulsivity and inattention problems associated with ADHD. There has been an assumption that there are no significant differences between auditory and visual CPTs, but research by Taylor (1994) found that "normal" college students were significantly more impulsive on auditory as compared to visual CPT tasks. The diagnostic validity of the IVA-QS, which includes both auditory and visual measures of impulsivity and inattention, was explored in this study. The purpose of this research was to determine whether this multi-sensory CPT's diagnostic accuracy was sufficient that it would be clinically valid to use routinely for individual assessments.

Thirty clients previously diagnosed by Dr. Sandford as having a primary diagnosis of ADHD were selected. His diagnosis was based on a full evaluation including comprehensive psychological testing and ADHD rating scales. These clients had volunteered and given permission to have their clinical test data used in this study. These data records were de-identified. The patient data used in this study were selected from patients seen between the years of 2001 and 2011. Fifteen of these individuals were male, and fifteen were female. The ages ranged from 6 to 55, and the mean age was 17.1. Most of these individuals were diagnosed solely as having ADHD, but a few had a secondary diagnosis of Cognitive Disorder, Not Otherwise Specified. In these cases they were found to have significant cognitive processing problems in addition to their attentional deficits. Thirty individuals from the normative database, matched by age and sex, were randomly selected for comparison to this group. None of these individuals from the normative group were identified as having been diagnosed

with ADHD, having ADHD-type symptoms, or having any other factors likely to impair their test functioning.

The ADHD diagnoses derived from the IVA-QS were compared to the clinical diagnosis, which was based on a full evaluation including diagnostic intake evaluation and comprehensive psychological testing, including ADHD rating scales. A comparison of the clinical diagnosis and the IVA-QS diagnosis for ADHD is presented below in Table 3-1a and 3-1b.

Table 2-1a. Comparison of Clinical Diagnosis and IVA-QS Diagnosis for ADHD

		Clinical Diagnosis		
		ADHD	No ADHD	Total
IVA-QS Test Diagnosis	ADHD	23	5	28
	No ADHD	7	25	32
	Total	30	30	60

Table 2-1b. Diagnostic Accuracy of the IVA-QS Test for ADHD

	Diagnostic Accuracy
Sensitivity	77%
Specificity	83%
Positive Predictive Power	82%
Negative Predictive Power	78%

Sensitivity is defined as the probability of the test identifying a positive result given that the individual has ADHD. Specificity is the probability of the test accurately identifying individuals who do not have ADHD. Positive Predictive Power is the percentage of individuals diagnosed by the test as having ADHD who were also clinically diagnosed.

Negative Predictive Power is the percentage of patients that were correctly diagnosed as not having ADHD.

The diagnostic accuracy of the IVA-QS test by itself was able to correctly identify about three out of four individuals who had been clinically diagnosed as having ADHD. The results show that the test misdiagnosed about two out of twelve individuals who did not have ADHD as having ADHD. Generally, rating scales are used by many clinicians to provide data in determining a diagnosis of ADHD. Thus, it is useful to compare the accuracy of the IVA-QS test to the accuracy of rating scales reported in the research literature in order to clinically evaluate the relative accuracy of the IVA-QS rating scales.

In evaluating individuals who sought treatment for clinical problems, Snyder, *et al.,* (2008) found that two different types of parent and teacher ADHD rating scales widely ranged in their diagnostic accuracy compared to clinicians' diagnoses which were based on a comprehensive, in-depth evaluation. The overall accuracy of ADHD rating scales for the clinical population in Snyder's study was ranged from 47% to 58%. This study's ADHD rating scale sensitivity in accurately diagnosing individuals with ADHD ranged from 38% to 78%. His results also found that the rating scales often mislabeled individuals as having ADHD who had either no diagnosis or another type of disorder. The low rates of specificity for this study ranged from 14-61%. A number of studies were also reviewed by Snyder that compared the accuracy of differentiating non-clinical populations from individuals diagnosed with ADHD, and the overall accuracy of these nine studies was higher, ranging from 55% to 79%.

The test results supported the clinical efficacy of the IVA-QS test by itself compared to ADHD rating scales. Based on the above review of the accuracy of ADHD rating scales, the IVA-QS test sensitivity of 77% is equivalent or better than the accuracy reported in most of the ADHD rating scale studies. The specificity of the IVA-QS (83%) is actually better than the highest accuracy rate for identifying individuals who do not have ADHD. In general, it has been noted that rating scales often have low rates of specificity in that many individuals without ADHD are misdiagnosed as having ADHD.

The rating scale diagnosis was also compared to the clinical diagnosis. In this case since rating scale data was not available for the matched normative sample group, only sensitivity and negative predictive power can be reported. The sensitivity of the rating scales alone was equal to the sensitivity for the IVA-QS CPT by itself, and the negative predictive power of the rating scales was slightly higher than that of the IVA-QS. (See Tables 3-2a and 3-2b presented below.)

Table 2-2a. Comparison of Clinical Diagnosis and ADHD Rating Scales

		Clinical Diagnosis		
		ADHD	No ADHD	Total
Rating Scale Diagnosis	ADHD	23	0	23
	No ADHD	7	30	37
	Total	30	30	60

Table 2-2b. Diagnostic Accuracy of ADHD Rating Scales

	Diagnostic Accuracy
Sensitivity	77%
Negative Predictive Power	81%

In this diagnostic validity study, the question is also addressed as to whether the IVA-QS CPT test results in combination with the rating scales may further help clinicians in accurately diagnosing ADHD. Consequently, the rating scale diagnosis for each client was determined by using the symptom cut-off guidelines for hyperactive/impulsive and inattentive symptoms appropriate to the client's age (American Psychiatric Association, 2013). Parent, teacher, and self-rating scales were combined. If any one of the available scales had ADHD symptoms above the cut-off for either hyperactive/impulsive or inattention symptoms, that positive symptom rating was used in formulating a diagnosis. In other words, if the parent rating scale identified six hyperactive/impulsive symptoms and the teacher identified six inattentive symptoms then the rating scales were interpreted as supporting the diagnosis of ADHD, Combined presentation.

The IVA-QS and rating scale data were combined in diagnosing ADHD. If either of these two methods supported a diagnosis of ADHD, then that diagnosis was assigned. There was no rating scale data for the normative population, so for the normative population, only the IVA-QS results were used in determining a diagnosis of ADHD. The results of the IVA-QS and rating scale diagnoses in comparison to the clinical diagnosis are presented below in Tables 3-3a and 3-3b.

Table 2-3a. Comparing ADHD Dx with the IVA-QS Test Results and Rating Scale Combined Dx

IVA-QS & Rating Scale Combined Diagnosis	Clinical Diagnosis: ADHD	Clinical Diagnosis: No ADHD	Total
ADHD	26	5	31
No ADHD	4	25	29
Total	30	30	60

Table 2-3b. Diagnostic Accuracy of the IVA-QS and Rating Scales Combined Dx for ADHD

	Diagnostic Accuracy
Sensitivity	87%
Specificity	83%
Positive Predictive Power	84%
Negative Predictive Power	86%

The combination of the IVA-QS test results and the rating scales increased the overall sensitivity by 10%. The specificity remained the same. The positive predictive power was essentially the same. The negative predictive power improved by 8%. Clinicians using solely this combined approach would be able to accurately identify individuals with ADHD in 13 out of 15 cases.

The value of including the rating scales in making a diagnosis of ADHD is likely to be due to the fact that they provide a measure of the occurrence of gross-motor hyperactivity that is not specifically identified by the IVA-QS test. In addition, the rating scales provide data relevant to the individual's functioning in both the home and school environments in respect to ADHD symptoms which may not manifest under the more controlled test conditions required for administering the IVA-QS. In contrast, the test results provide the clinician with the opportunity to objectively measure clients' mental processing speed and its variability, attentional functioning, and impulsive responses that may be aspects of ADHD that are difficult for raters to accurately identify in the work and school environments. Separately, the rating scale and the IVA-QS are equivalent in respect to their accuracy in diagnosing ADHD, but in combination, they were found to be more accurate in this study. These findings support combining the IVA-QS test results with rating scales in helping clinicians to make more accurate diagnoses of ADHD.

Test-Retest Reliability Analysis

Reliability refers to the consistency of test score performance for repeated testing by an individual under similar conditions. This type of reliability test provided an index about the stability of IVA-QS test scores over time (Anastasi, 1988) that allows health professionals to be confident that the changes observed in scores reflect differences in a person's performance and are not solely due to random errors. If the IVA-QS test results are found to be consistent over time, then they can be practically used in clinical diagnostic decision making and in the evaluation of medication and/or treatment effects. The purpose of this study was to determine the test-retest reliability of the 8 global scale scores. These global scales are used in evaluating overall attention problems and in helping clinicians make diagnostic decisions. The global quotient scale scores are based on the raw scores of selected relevant scales and statistically derived. The global scale scores were selected to determine test reliability, because they are the best measure of an individual’s overall performance.

A total of non-clinical 20 volunteers without identified problems of neurological, current psychological, learning, attention, or self-control problems were administered the IVA-QS on two separate occasions. These data records were de-identified. The age range of volunteers was 6 to 68 years old, and the mean age was 26.5 years. Forty-five percent of these individuals were females and 65% were males. The test data analyzed was collected for educational purposes from several different settings.

The test was administered as described in the manual. The first and second tests were given 53 days apart on average. All participants were given this opportunity to learn how to click the mouse correctly. Next, each participant completed a 32 item practice session to learn with the examiner's help, if necessary, how to respond correctly when both targets (1's) and foils (2's) were presented in a mixed-up order. If during the practice session the test taker had problems doing the test task correctly, then the examiner could temporarily stop the practice training and explain the four simple test rules which were:

1. Click when you see a "one"
2. Click when you hear a "one"
3. Don't click when you see a "two"
4. Don't click when you hear a "two"

Once the main part of the IVA-QS test began, no further instructions could be given, except to redirect the test taker if he or she removed his/her finger from the correct mouse button. At the end of the test scores were then automatically saved by the computer for later analysis.

Two tables are presented below which summarize the data of this reliability analysis. Statistical analyses using an ANOVA comparing each scale were completed to determine whether test scores were similar for each individual for both testing sessions and whether the scores changed due to any practice, fatigue, or motivational effects.

The IVA-QS was found to be a test that is robustly stable over time, because retesting was not found to not change the test scores in any significant way. The results shown in Table 2-1 reveal that there are no significant differences for any of the global IVA-QS test scores when individuals were retested. In Table 2-2, the correlations for the global test quotient scores were all significant and positive. The correlation effect size was very strong for the majority of these scales. Consequently, the IVA-QS can be used with confidence to measure any treatment or medication effects by retesting individuals.

It would be valuable in the future to also analyze and compare test performance on and off of medications for those identified with attention or hyperactive/impulsive symptoms. This could help evaluate the effectiveness of the IVA-QS in assessing attention differences or response control differences due to medication interventions.

Table 3-1. Comparison of Mean Global Quotient Scale Scores for Test-Retest

Global Scale	Mean Q-Score		p-value
	Initial Test	Retest	
Full Scale Response Control (FRCQ)	103.8	103.3	n.s.
Auditory Response Control (ARCQ)	105.1	104.6	n.s.
Visual Response Control (VRCQ)	101.4	101.4	n.s.
Full-Scale Attention (FAQ)	109.2	113.3	n.s.
Auditory Attention (AAQ)	111.3	111.0	n.s.
Visual Attention (VAQ)	104.5	111.8	n.s.
Sustained Auditory Attention (SAAQ)	109.1	109.1	n.s.
Sustained Visual Attention (SVAQ)	105.4	112.3	n.s.

Table 3-2. Test-Retest Correlations for IVA-QS Global Scale Scores

Global Scale	r	p-value	Effect Size
Full-Scale Response Control (FRCQ)	0.56	< 0.01	Moderate
Auditory Response Control (ARCQ)	0.60	< 0.01	Strong
Visual Response Control (VRCQ)	0.46	< 0.05	Moderate
Full-Scale Attention (FAQ)	0.86	< 0.001	Very Strong
Auditory Attention (AAQ)	0.86	< 0.001	Very Strong
Visual Attention (VAQ)	0.82	< 0.001	Very Strong
Sustained Auditory Attention (SAAQ)	0.90	< 0.001	Very Strong
Sustained Visual Attention (SVAQ)	0.87	< 0.001	Very Strong