FDA Executive Summary

Prepared for the June 3-4, 2021, Meeting of the Neurological Devices Advisory Panel

Classification of Attention Task Performance Recorders

Product Code: LQD

Table of Contents

List of Tables

List of Figures

1. Introduction

Per Section 513(b) of the Food, Drug, and Cosmetic Act (the Act), the Food and Drug Administration (FDA) is convening the Neurological Devices Advisory Panel (the Panel) for the purpose of obtaining recommendations regarding the classification of attention task performance recorder, a pre-amendments device type which remains unclassified. Specifically, the FDA will ask the Panel to provide recommendations regarding the regulatory classification of attention task performance recorder under product code "LQD." The device names and associated product codes are developed by the Center for Devices and Radiological Health (CDRH) in order to identify the generic category of a device for FDA. While most of these product codes are associated with a device classification regulation, some product codes, including "LQD," remain unclassified.

FDA is holding this Panel meeting to obtain input on the risks to health and benefits of the attention task performance recorders under product code "LQD." The Panel will discuss whether the attention task performance recorders under product code "LQD" should be classified into Class II (subject to General and Special Controls).

1.1 Current Regulatory Pathways

Attention task performance recorders are a pre-amendments, unclassified device type. This means that this device type was marketed prior to the Medical Device Amendments of 1976 but was not classified by the original classification panels. Currently these devices are being regulated through the 510(k) pathway and are cleared for marketing if their intended use and technological characteristics are "substantially equivalent" to a legally marketed predicate device. Since these devices are unclassified, there is no regulation associated with the product code.

1.2 Device Description

An attention task performance recorder is intended to measure reaction time (RT) in response to attention tasks. An attention task performance recorder may or may not be used to aid in the assessment or diagnosis of specific clinical conditions, most specifically attention deficit hyperactivity disorder (ADHD). For general assessment of RT, the device may provide measures of both the speed of responding to stimuli and how accurately patients respond to stimuli without specific use and without providing clinical context regarding a specific disease or condition (e.g., comparison to a normative database for a clinical condition). For the assessment of specific clinical conditions (e.g., ADHD), the device may additionally provide information regarding correlation with known neuropsychometric tests or aspects of cognition related to the condition of interest.

In terms of technological characteristics, the devices are typically software-based, with a test or evaluation being manually administered by a clinical end user for assessment of the symptom(s) of interest. For example, the T.O.V.A., QbTest and QbCheck devices, which have been cleared under product code LQD (see also Table 1 below), are software-based tests intended to provide objective measures of the core symptoms of ADHD (activity, attention and impulsivity) using a motion tracker. Tests supported by the attention task performance recorders can be administered via software loaded on a host computer, with clinicians being able to view results alongside patient history to inform further decision making. Another example of an attention task

performance recorder device is the Gordon Diagnostics System, which additionally includes a hardware unit for task administration.

2. Regulatory History

The Gordon Diagnostic System Model I was the first device cleared under product code LQD on June 2, 1986. The FDA determined that the Gordon Diagnostic System Model I was substantially equivalent to pre-amendments, unclassified attention task performance recorders. Please refer to Table 1 for a listing of the manufacturers, device names, and associated 510(k) submission numbers for cleared attention task performance recorders under product code "LQD":

Table 1: 510(k) Clearances for Attention Task Performance Recorders under Product Code "LQD"

510(k) Number	Trade Name	Sponsor
K854903	Gordon Diagnostic System Model I	Clinical Diagnostics, Inc.
K861304	Fagan Test Machine for Infant Intelligence	InfanTest Corp.
K911938	DynaVision 2000	Performance Enterprises
K020800	OPTAx System	OPTAx Systems Inc.
K040894	QbTest	Qbtech AB
K122149	QbTest	Qbtech AB
K133382	QbTest	Qbtech AB
K143468	QbCheck	QbTech AB
K141865	DANA	AnthroTronix, Inc.
K170082	Test of Variables of Attention (T.O.V.A.), version 9.0	The TOVA Company
K173915	Test of Variables of Attention (T.O.V.A.), version 9.0	The TOVA Company

3. Indications for Use

The Indications for Use (IFU) statement identifies the condition and patient population for which a device should be appropriately used. The IFU statements for the cleared devices under product code LQD are specified in Table 2 below. All of the devices are cleared for prescription use only.

Table 2: Indications for Use of 510(k)-Cleared Devices under Product Code "LQD"

510(k) Number	Indications for Use
K854903	Portable electronic device designed to assess deficits in attention and impulse control in children. It has been developed for use by clinicians as an aid in the diagnosis of attention deficit disorders as well as some forms of learning disabilities.
K861304	The Fagan Test of Infant Intelligence is intended to screen high risk babies (those confined to NICU's or born to diabetic mothers for example) from those most likely to suffer later cognitive deficit. The test consists of

510(k) Number	Indications for Use
	the presentation of a precisely ordered series of pictures to the infant at different ages and recording the amount of time the infant spends looking at each image.
K911938	Measurement of reaction time. Tests visual reaction speed, physical response speed, and overall motor response time.
K020800	The OPTAx System provides clinicians with objective measurements of hyperactivity, impulsivity and inattention to aid in the clinical assessment of ADHD. OPTAx results should be interpreted only by qualified professionals.
K040894	QbTest provides clinicians with objective measurements of hyperactivity, impulsivity, and inattention to aid in the clinical assessment of ADHD. QbTest results should be interpreted only by qualified professionals.
K122149	QbTest is indicated to be used to aid in the clinical assessment of ADHD. QbTest results should be interpreted by qualified health care professionals only.
K133382	QbTest provides clinicians with objective measurements of hyperactivity, impulsivity and inattention to aid in the clinical assessment of ADHD (Attention Deficit Hyperactivity Disorder) and in the evaluation of treatment interventions in patients with ADHD. QbTest results should be interpreted only by qualified professionals.
K143468	QbCheck provides health care professionals with objective measurements of hyperactivity, impulsivity, and inattention to aid in the clinical assessment of ADHD and in the evaluation of treatment interventions in patients with ADHD. QbCheck results should be interpreted only by qualified health care professionals.
K141865	DANA provides clinicians with objective measurements of reaction time (speed and accuracy) to aid in the assessment of an individual's medical or psychological state. Factors that may affect the measurement of reaction time include, but are not limited to concussion, head injury, insomnia, post traumatic stress disorder (PTSD), depression, attention deficit hyperactivity disorder (ADHD), memory impairment, dementia, delirium, prescription and non-prescription medication, some nutritional supplements, as well as a variety of psychological states (e.g. fatigue and stress). DANA also delivers and scores standardized psychological questionnaires. DANA results should be interpreted only by qualified professionals.
K170082	The Test of Variables of Attention (T.O.V.A.) provides healthcare professionals with objective measurements of attention and inhibitory control, which aid in the assessment of attention deficits, including attention-deficit/hyperactivity disorder (ADHD). T.O.V.A. results should only be interpreted by qualified healthcare professionals.
K173915	The Test of Variables of Attention (T.O.V.A.) provides healthcare professionals with objective measurements of attention and inhibitory control. The visual T.O.V.A. aids in the assessment of, and evaluation of

510(k) Number	Indications for Use
	treatment for, attention deficits, including attention-deficit/hyperactivity disorder (ADHD). The auditory T.O.V.A. aids in the assessment of attention deficits, including ADHD. T.O.V.A. results should only be interpreted by qualified professionals.

4. Clinical Background

4.1 Disease Characteristics

ADHD is a common mental health disorder of childhood affecting approximately 4–8% of school-aged children[1]. This neurodevelopmental disorder is characterized by three core symptom domains: inattention, hyperactivity and impulsivity. It is normal for children to have trouble focusing and behaving at one time or another. However, children with ADHD do not just grow out of these behaviors. The symptoms continue, can be severe, and can cause difficulty at school, at home, or with friends. A child with ADHD might: daydream frequently, forget or lose items frequently, squirm or fidget, talk too much, make careless mistakes or take unnecessary risks, have a hard time resisting temptation, have trouble taking turns, or have difficulty getting along with others.

There are three different types of ADHD, depending on which types of symptoms are strongest in the individual[2]:

- Predominantly Inattentive Presentation: It is hard for the individual to organize or finish a task, to pay attention to details, or to follow instructions or conversations. The person is easily distracted or forgets details of daily routines.
- Predominantly Hyperactive-Impulsive Presentation: The person fidgets and talks frequently. It is hard to sit still for long (e.g., for a meal or while doing homework). Smaller children may run, jump or climb constantly. The individual feels restless and has trouble with impulsivity. Someone who is impulsive may interrupt others a lot, grab things from people, or speak at inappropriate times. It is hard for the person to wait their turn or listen to directions. A person with impulsiveness may have more accidents and injuries than others.
- Combined Presentation: Symptoms of the above two types are equally present in the person.

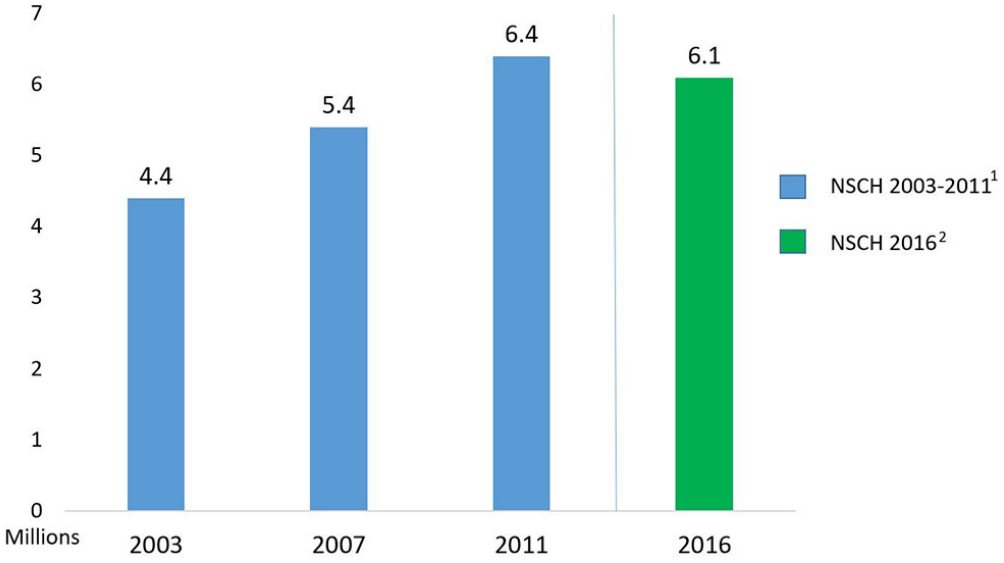
Because symptoms can change over time, the presentation may change over time as well.

Millions of Unites States (US) children have been diagnosed with ADHD (see Figure 1). The estimated number of children ever diagnosed with ADHD, according to a national 2016 parent survey[1], is 6.1 million (9.4%). This number includes:

- 388,000 children aged 2–5 years;
- 4 million children aged 6–11 years; and
- 3 million children aged 12–17 years.

Boys are more likely to be diagnosed with ADHD than girls (12.9% compared to 5.6%)[1].

Figure 1: Estimated number of US children who ever had a diagnosis of ADHD.



About this chart:

1 NSCH 2003-2011: National Survey of Children's Health, telephone survey data; estimate includes children 4-17 years of age.
2 NSCH 2016: Redesigned as an online and mail survey, estimate includes children 2-17 years of age.

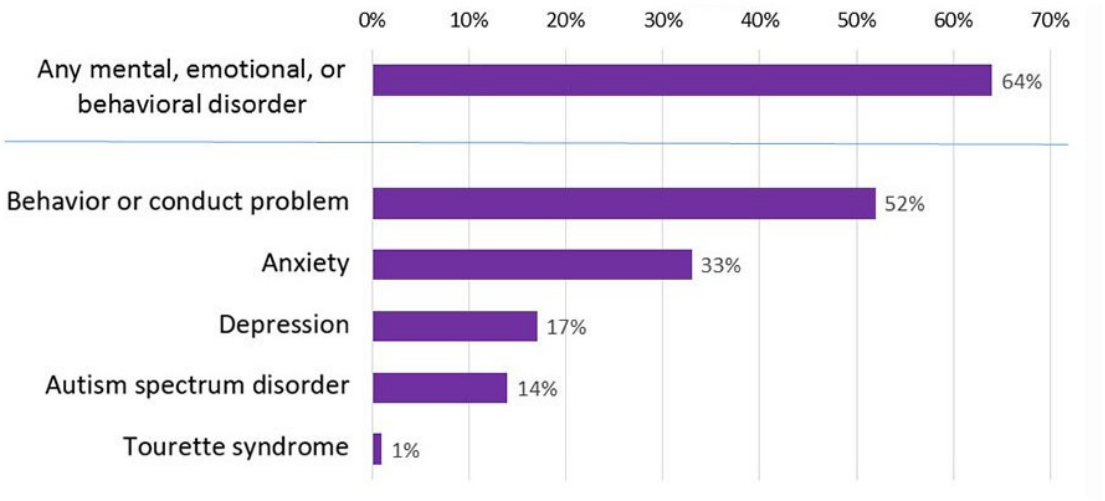
Because the 2016 NSCH survey used different methods, estimates are not directly comparable with estimates based on previous NSCH data. Because of an increased focus on ADHD in younger children, age ranges were expanded to include children 2-17 years of age

According to a national 2016 parent survey, 6 in 10 children with ADHD had at least one other mental, emotional, or behavioral disorder (see Figure 2)[1]:

- About 5 in 10 children with ADHD had a behavior or conduct problem.
- About 3 in 10 children with ADHD had anxiety.

Other conditions affecting children with ADHD include depression, autism spectrum disorder, and Tourette syndrome.

Figure 2: Percentage of Children Diagnosed with ADHD with Additional Disorders



In practice, delivery and quality of care for those diagnosed with ADHD are patchy, with little consistency in assessment, diagnosis or management[3,4]. It has been suggested that ADHD is 'symptom complex', stemming from multiple causes, such as genetics, biological and psychosocial influences, resulting in a range of presenting behaviors[3-5]. Given the variation in causes and behavioral consequences of ADHD, there is no single test used to diagnose the disorder, and the clinician's judgment is currently the most widely accepted method of assessment. For the clinician to determine a diagnosis of ADHD, they will generally gather information from the parents, teachers (and the child themselves where age appropriate), make clinical observations, conduct school observations, and may use tests of behavior and neuropsychological functioning. However, there is a paucity of clinical guidance on which combination of measures should be used in the diagnostic assessment of ADHD.

Furthermore, this approach is heavily reliant on subjective measures, which can lead to discrepancies in the diagnosis of ADHD[4], and the process of interview and data collection is lengthy and difficult to conduct in real-world settings. Additionally, once on medication, monitoring may not be adequate or frequent enough to detect early non- or sub-optimal response[3,5]. Objective measures have the potential to augment and streamline current practice in order to shorten assessment time, increase diagnostic accuracy, reduce delays in treatment, and optimize treatment response. Continuous performance tests (CPTs), such as attention task performance recorders, are objective neuropsychological tests that measure the individual's attention and impulsivity in a sustained task and can be used alongside clinical inquiry as part of the diagnostic procedure.

4.2 Diagnosis

The American Academy of Pediatrics (AAP) guidelines for diagnosis and evaluation of ADHD recommend that primary care providers complete these steps:

- Evaluate children and adolescents ages 4 to 18 years for ADHD if they are having academic or behavioral problems and show inattention, hyperactivity, or impulsivity.

- Get reports on the child's symptoms from parents or guardians, school staff, and mental health workers involved with their care, and get information from the child or adolescent as well.
- Use rating scales and other sources to document the symptoms and ensure that the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria have been met.
- Rule out any other possible conditions that can cause similar symptoms.
- Screen for other conditions that might coexist with ADHD, including emotional or behavioral disorders (such as anxiety, depression, and behavior problems), developmental disorders (such as learning and language disorders or autism spectrum disorder (ASD)), and physical conditions (such as tics, sleep disorders, or apnea).
- **Refer children to a specialist if they detect coexisting conditions that they are not experienced in treating or diagnosing.**

4.4 Risks

FDA has identified the following probable risks to health associated with attention task performance recorders intended to 1) measure reaction time and associated patient performance in response to attention tasks and 2) aid in assessment or diagnosis of specific disease or conditions.

Table 3: Risks to Health and Descriptions/Examples for Attention Task Performance Recorders Intended to Measure Reaction Time and Associated Patient Performance in Response to Attention Tasks Only, Without Aiding in Assessment or Diagnosis

Identified Risk	Description/Examples
Patient discomfort (e.g., visual or mental fatigue)	• Use of the devices can cause patient discomfort, such as visual or mental fatigue.
Incorrect or inaccurate measurements of reaction time or other attention tasks	• Use of the devices can result in incorrect or inaccurate measurements of reaction time or other attention tasks based on associated patient performance

Table 4: Risks to Health and Descriptions/Examples for Attention Task Performance Recorders Intended to Aid in Assessment or Diagnosis of Specific Diseases or Conditions

Identified Risk	Description/Examples
Patient discomfort (e.g., visual or mental fatigue)	• Use of the devices can cause patient discomfort, such as visual or mental fatigue.
Incorrect or inaccurate results leading to inaccurate assessment or delayed diagnosis, both of which could result in inappropriate therapy or delay in treatment	• A false positive result means that the device indicates the patient has the clinical condition or disease of interest, such as ADHD or be at risk of cognitive impairment, when in fact none is present. • A false negative result means that the device indicates the patient does not have the clinical condition or disease of interest,

Identified Risk	Description/Examples
	such as ADHD or be at risk of cognitive impairment, when in fact the clinical condition or disease is present.

The Panel will be asked whether this list is a complete and accurate list of the risks to health presented by attention task performance recorders under product code "LQD" and whether any other risks should be included in the overall risk assessment of the device type.

5. Literature Review

5.1 Methods

A systematic literature review (SLR) was conducted in an effort to gather any published information regarding the safety and effectiveness, and specifically validity and reliability, of attention task performance recorders that are regulated under product code "LQD". The search was limited to human clinical studies published in English language and with publication dates between January 1, 2010, and December 31, 2020. Online literature searches were performed in two electronic databases (MEDLINE and Embase) using search terms limited to attention task performance recorders that are cleared for market distribution in the US (listed in Table 1), and keywords from their IFU (see Appendix A for detailed methodology).

After the results from each set of search terms were combined (MEDLINE n=177, Embase n=169) and 115 duplicate references were removed, a total of 231 articles remained. An additional 10 studies were identified through other sources (i.e., references captured in our literature search or in 510(k) submissions of attention task performance recorders). Figure 3 in Appendix A presents the screening process and selection of references for inclusion in this SLR. Studies were included if they evaluated attention task performance recorders within their IFUs, their validity and reliability, and if the studies used one of the following rating scales when evaluating the validity of the recorder: (1) the Attention Deficit Hyperactivity Disorder Rating Scale (ADHD-RS); (2) the Bayley Scales Infant Development (BSID); (3) the Conners Comprehensive Behavior Rating Scales (Conners); (4) Swanson, Nolan and Pelham Teacher and Parent Rating Scale (SNAP), and (5) Vanderbilt ADHD Diagnostic Rating Scales (Vanderbilt). Studies were excluded if they were conducted outside the US (unless only outside the US studies were identified for the specific device). Based on the review of 241 titles and abstracts, 101 references were excluded (see Appendix A, Figure 3 for details on exclusion reasons). The remaining 140 references underwent full text review, which resulted in the exclusion of an additional 98. This resulted in 42 references identified as eligible to be included in this SLR. These references report on 41 studies.

5.2 Results

Among the 41 studies included in this evidence assessment, there were 2 SLRs[6,7], 1 meta-analysis[8], 11 randomized control trials (RCTs)[9-21], 5 experimental studies[22-26], 5 test-retest reliability studies[27-31], 4 validation studies[32-35], 5 cohort studies[36-40], 4 case-control studies[8,41-43], and 4 cross-sectional studies[44-47].

The search methodology captured references presenting results on attention task performance recorders with the following intended uses, to:

- provide objective measurements of reaction time (RT), assessed by speed and accuracy, to aid in the clinical assessment of neurological status (Defense Automated Neurobehavioral Assessment (DANA) and Dynavision 2000);
- provide measures of novelty preference as predictive of later cognitive functioning in infants at high risk of intellectual disability (Fagan Test Machine for Infant Intelligence (Fagan Test));
- aid in the clinical assessment of attention deficits, including ADHD, by providing objective measurements of impulsivity, inattention (Gordon Diagnostic System (GDS), Test of Variables of Attention (T.O.V.A.)), and hyperactivity (OPTAx System (OPTAx), QbTest, and QbCheck); or
- aid in the evaluation of treatments for patients with ADHD (QbTest, QbCheck, and T.O.V.A.).

The search also captured studies on the OPTAx, QbTest, and QbCheck and reported on objective measurements of hyperactivity. Outside the US studies evaluating QbTest and QbCheck (n=16) were included because no US studies were identified for these devices[14-21,24,26,31-35,38-40]. Thirteen (13) studies had results on test-retest reliability[9,27-31]; internal reliability[24,30,34,35]; construct validity, measured as convergent validity[8,19,21,31,34,35,38] and discriminant/divergent validity[34,35]; and/or concurrent validity[9,31,34]. The studies identified included sample sizes ranging between 13 and 1,200 subjects and evaluated subjects as young as infants and up to 79 years old.

5.3 Adverse Events Associated with Attention Task Performance Recorders

The search methodology did not identify literature reporting on adverse events related with the use of the attention task performance recorders themselves.

5.4 Effectiveness Associated with Attention Task Performance Recorders

The search identified 13 studies evaluating RT[22,23,27-30,37,41-46], 25 studies evaluating their use to aid in the clinical assessment of ADHD (n=9 studies[8,9,31-35,39,47] and 1 meta-analysis[8]) or in the evaluation of treatment interventions (n=15 studies[10-21,24-26,38,40]) in patients with ADHD, and 2 studies evaluating the use of the Fagan Test as a cognitive screening tool[7,36]. In addition, we identified 1 SLR evaluating the clinical utility of attention task performance recorders for diagnosing and monitoring ADHD in children[6]. Literature review results for the effectiveness associated with attention task performance recorders are summarized below according to their indications.

Measurement of Reaction Time

There were 8 studies that used DANA to measure RT (2 experimental studies[22,23], 1 cohort study[37], 1 case-control study[41], 2 cross-sectional studies[44,45], and 2 test-retest reliability studies[29,30]). Most of the studies (n=6) were conducted with military or law enforcement personnel. Dynavision 2000 was used in 5 studies (2 case-control studies[42,43], 1 cross-sectional study[46], and 2 test-retest reliability studies[27,28]).

Studies have used the DANA to measure RT and response inhibition (Go/No-Go (GNG) test) in different study populations and environments. In a case-control study of 29 adolescents, those adolescents that had a concussion had a statistically significant lower throughput score (rate of correct responses per minute) for simple RT (SRT) and response inhibition than controls[41]. Procedural RT (PRT) scores were not statistically significantly different between cases and

controls. In a cross-sectional study of 646 subjects, neither single previous concussion nor recency of concussion were associated with neurocognitive outcomes after adjustment for post-traumatic stress disorder (PTSD), depression, and deployment experience. However, those reporting 3 or more lifetime concussions performed worse on neurocognitive tasks requiring simple attention and simple discrimination skills, including SRT, response inhibition, and code substitution (CDS)[45]. Servatius et al.[39] conducted a cross-sectional study in 241 US Coast Guard personnel to assess stress-related mental health symptoms of PTSD and major depressive disorder (MDD), personality, and neurocognitive function. Impaired neurocognitive performance was concentrated among those with both PTSD and MDD. The poor performance was mostly driven by deficits in SRT and response inhibition (GNG).

Sub-concussive blast effects on neurocognitive performance were also observed in an experimental study involving 202 subjects from the US Army exposed to explosive blast overpressure[22]. The "High" exposure group had a statistically significant slower mean PRT compared to the "Low" exposure group within 5 minutes of exposure and at the end of the day. Response inhibition within 5 minutes of exposure was also slower in the "High" exposure group ($p < 0.05$). No statistically significant difference was observed on SRT between exposure groups. Blast overpressure effects from 50 caliber weapon usage was also evaluated in 20 military and law enforcement male professionals over a 3-day training period[37]. DANA baseline test results for each training day were compared to those recorded after and at the end of the day. Only PRT and GNG tests showed statistically significant slowed reaction times over the progression on days 1 and 2, but not day 3.

Altitude-related performance was evaluated in an experimental study where 21 subjects had neurocognitive assessments at sea level, immediately after ascending to an elevation over 5000 m, and following 16 days of acclimatization to this high altitude[23]. Only a marked decrease in the differences in SRTs (dSRT) score was observed after ascent from sea level ($p < 0.005$). dSRT score also showed the remission of impairment after acclimatization to high altitude (i.e., dSRT score was similar to those observed at sea level).

Lathan et al.[30] conducted a test-retest reliability study to test the DANA in 244 active duty US service members in 5 environments (desert, jungle, mountain, arctic, and shipboard). The authors report that the DANA performed well in these environments and combined the data since the data sets were not statistically significantly different. They calculated the intra-class correlation coefficients (ICCs) across 10 administrations over 2 days. For SRT, PRT, GNG, spatial processing (SPD), and CDS, good to excellent reliability was achieved with ICC values of 0.95, 0.91, 0.95, 0.89, and 0.88, respectively. The ICC was moderate for code substitution recall (CDD; r=0.54). This study also evaluated the DANA's internal reliability using split-half correlations (r). Good to excellent and statistically significant correlations ($r \geq 0.76$) for SRT, PRT, GNG, CDS, SPD, and CDD were observed. Another study evaluated the test-retest reliability and stability of performance of the DANA within different environments (thermoneutral, simulated (cold, hot, and humid)) in 16 subjects. There were no statistically significant differences observed between the different environments. The mean test-retest reliability across subtests was lowest in the cold ($ICC_{2,1}=0.53$) and highest in the hot environment ($ICC_{2,1}=0.77$)[29]. The range of ICCs within each subset were as follows: CDS ($ICC_{2,1}$ range: 0.60-0.91), GNG ($ICC_{2,1}$ range: 0.54-0.76), PRT ($ICC_{2,1}$ range: 0.54-0.89), SPD ($ICC_{2,1}$ range: 0.47-0.84), SRT ($ICC_{2,1}$ range: 0.5-0.85), and SRT2 ($ICC_{2,1}$ range: 0.46-0.90). No practice effects were observed in this study. Overall, test-retest

reliability results from these studies were similar, including both reporting the lowest reliability for CDD. Lathan et al. results tend to be on the higher end of the ICCs range; however, further comparisons are limited since there were no ICCs presented by environment. It is also not clear in Lathan et al.'s study if they used the recommended model to calculate ICCs to evaluate test-retest reliability (single rater, absolute agreement, 2-way mixed effects analysis of variance model ($ICC_{2,1}$ formula)[48,49]) or the formula where the relationship is defined as consistency instead of absolute agreement ($ICC_{3,1}$)[48]. Koo and Li showed that unless the data sets are identical, $ICC_{3,1}$ generally gives a larger ICC than $ICC_{2,1}$[48]. Despite these limitations, performance on the DANA seems to be reliable in different environments.

Two (2) studies used Dynavision 2000 to evaluate the effect of concussion on RT. In a case-control study of 23 adults with concussion and 30 adults without concussion, mean central and peripheral vision RTs were statistically significantly longer for cases than for controls[42]. Another case-control study of 13 cases (collegiate student-athletes who have returned to baseline on clinical concussion assessments) and 13 controls did not show differences in visual motor coordination ((VMC) A* exam, simple visual RT, and simple visual movement time) between cases and controls[43]. The authors note that there may be a practice effect since both groups significantly improved their score and RT between sessions despite subjects completing the manufacturer-recommended warm-up designed to reduce practice effect. They also note that it may be "either there are no lingering deficits in VMC post-concussion in this population or the Dynavision was not sensitive enough to identify the lingering deficits".

The test-retest reliability of the Dynavision was evaluated in healthy adults in 2 studies[27,28]. In the first study, Wells et al.[28] asked 42 recreational active young adults to complete 6 trials of 3 RT tasks of increasing complexity (Choice RT, Mode A, and Mode B) in 2 sessions separated by 48 hours[28]. Choice RT measures visual and motor RT (VRT and MRT) to a visual stimulus with the dominant hand. Mode A measures the ability to react to a stimulus as it changed positions on the board at random locations within 60 seconds using both hands. Mode B adds the complexity of participants to verbally recite a five-digit number presented on the LCD screen. Moderate to good reliability was shown for VRT ($ICC_{2,1}$=0.84), MRT, and reactive ability in both Mode A and Mode B tasks ($ICC_{2,1}$=0.63-0.75). The results of this study also showed that 1 practice trial was needed for the CRT task and 3 for reactive RT tasks. In the second study, Picha et al. evaluated the test-retest reliability of 5 new 60-seconds Dynavision RT protocols at different time intervals (1 hour and 14 days apart)[27] in 30 healthy young adults. The first 3 protocols (Speed, Simple, and Moderate) were similar to the ones used in the study by Wells et al.[28] The Speed protocol is identical to Mode A and the Simple protocol lasts 15 seconds longer than the CRT task. The Moderate protocol and Mode B task are similar except that the Moderate protocol challenges participants with solving simple math equations instead of verbally reading a five-digit number. The last 2 protocols were more challenging. The Difficult protocol requires the participant to read a passage aloud as it is scrolled across the screen while extinguishing the lights and the GNG adds green lights that participants were instructed to avoid hitting while continuing to hit the red lights. All protocols had good reliability between the 3 sessions ($ICCs_{3,1}$=0.75-0.90). ICCs for the 3 similar protocols were higher in this study compared to Wells et al.'s ($ICCs_{2,1}$=0.63-0.72). The authors note that this difference may be due to additional practice trials or a potential training effect. An additional explanation for the larger ICCs reported could be that the ICCs were calculated using the $ICC_{3,1}$ model which results in higher ICCs[48].

Blackwell et al. [46] conducted a cross-sectional study in 300 healthy adults (ages 18 to 80 years old) to provide normative data for the Dynavision D2 for physical response speed. Statistically significant differences were observed in physical response speed between men and women and between the different age groups. Women in all age categories were slower than men. Physical response speed increased with age in both sexes.

In summary, the DANA was able to detect differences in measures of SRT and response inhibition between subjects diagnosed with concussions and controls. Among subjects exposed to sub-concussive blast, PRT was the most sensitive to identifying performance changes. DANA was also able to capture differences in SRTs when subjects experienced hypoxic conditions at high altitude. The DANA was also shown to perform reliably across different type of environments and to have good internal reliability, but these results are based on 2 studies (n=16-224). Based on the results from these studies, the data seem to suggest that the DANA is able to measure RT reliably and there is greater uncertainty for the Dynavision to reliably measure RT.

Cognitive Screening Tool

The Fagan Test was developed for the early detection of later intellectual disability. It provides a novelty preference score, which is the proportion of time the infant spends looking at a novel picture in relation to time looking at a familiar picture, presented in 3 categories: low risk, suspect, or at risk. One (1) of the 2 publications identified is a SLR that examined the utility of the Fagan Test in infants with cerebral palsy or motor impairment[7]. Morgan et al.[7] identified 7 relevant articles (6 studies) published between 1986 and 2006. The studies had small samples sizes ranging from 18 to 196 infants and a third were conducted in high-risk infants for later cognitive impairment, the population the Fagan Test was developed for. One (1) of the studies screened 62 infants at risk for later intellectual disability and assessed their cognitive development at 3 years of age. The Fagan Test was administered at least twice between 3 and 7 months of age[50]. The sensitivity (sens) and specificity (spf) of the Fagan Test for identifying infants with delayed cognitive development at 3 years were 75% (6/8) and 91% (49/54), respectively. The positive predictive value (PPV; i.e., the probability of an infant with a positive Fagan Test truly has delayed cognitive development at 3 years) was 55% (6/11) and the negative predictive value (NPV) was 96% (49/51), respectively in this study population that had a prevalence of 13% delayed cognitive development. Morgan et al. noted that this study had the limitation of using a criterion that cannot be considered an adequate "gold standard" for measuring intelligence quotient (IQ). In the other study, the Fagan Tests administered in 18 Italian infants at 9 and 12 months was found statistically significantly correlated with the neurodevelopmental outcome (Griffiths score) at 2 years (9m: sens:100%, spf:68%, p=0.016; 12m: sens:50%, spf: 87%, p=0.049)[51]. A limitation of this study is its small sample size. Despite these limitations, Morgan et al. concluded that the Fagan Test have predictive utility in this population. Their recommendation was to "probably" use it in clinical practice and research for predicting future intelligence. The studies evaluating the Fagan Test had methodological issues and conflicting findings for discriminating normal from abnormal cognitive skills. Their recommendation was to "probably don't use" the Fagan Test for discriminating normal from abnormal cognitive skills. The other study using the Fagan Test identified was a cohort study of 299 infants with and without HHV-6 infection[36]. Fagan Test assessments were done at 4 and 6 months of age and BSID-II Mental Developmental Index (MDI) assessments at 12 months. The 12-month follow-up rate was much lower in the HHV-6 infected group (68% (39/57)) compared to the non-infected group (82% (199/242)). No statistically significant differences between infants with and without congenital

HHV-6 infection were identified for Fagan Tests of novelty preference, mean fixation duration for familiarization trials, or mean fixation duration for novelty preference trials. However, infants with HHV-6 congenital infection had a lower mean (±SD) BSID-II MDI score (103.4±8.9) at age 12 months compared to the matched comparison group that had a mean score of 105.4±12.4. After controlling for gestational age, type of feeding, and age at test covariates, HHV-6 congenital infection was associated with lower scores on the BSID-II MDI at 12 months of age (mean difference: 4.3 [95% confidence interval (CI): 0.4, 8.1]; p=0.03) compared with infants without HHV-6 congenital infection.

In addition to the limited number of studies identified for this device, the studies identified in Morgan et al.'s SLR had study design issues. For example, two thirds of the studies (4/6) used the Fagan Test in normally developing infants and not the intended population (i.e., infants suspected to be at risk for later cognitive deficit). It is important to note that the PPV of a test increases as the prevalence of the disease/outcome increases. When the Fagan Test is used in a group of typically developing infants, a lower proportion of infants classified "at risk" will actually develop intellectual disability and there will be more false positives (infants labeled at risk that will not likely develop an intellectual disability), since the PPV will be low. Potential negative implications for false positives could include parental anxiety and limiting the infants' activities thinking that he/she has an intellectual disability, among other consequences. Another issue was that the 2 studies conducted among high-risk infants used criteria for predictive validity that were not considered gold standard or had a small sample size (n=18).

Aid in the Clinical Assessment of ADHD

A SLR[6], a meta-analysis[8], and 12 studies[8,9,19,21,24,31-35,38,39] evaluating the validity and/or reliability of attention task performance recorders used to aid in the clinical assessment of ADHD were identified. One (1) study evaluated the long-term temporal stability of measured inattention and impulsivity using GDS[47]. Most studies identified evaluated QbTest (n=10)[8,19,21,24,32-35,38,39].

Hall et al.[6] conducted a SLR to provide an overview of the evidence for attention task performance recorders that have been used for aiding the clinical diagnostic and medication monitoring for children and young people (up to 18 years old) with ADHD. Their search included studies published up to June 2015 without geographical restrictions. Their SLR identified 19 studies with results on the clinical utility to aid ADHD assessment for GDS (n=12), T.O.V.A. (n=6), and QbTest (n=1).

A study using the GDS investigated the stability of measured inattention and impulsivity in 562 children diagnosed with ADHD-Combined (ADHD-C) type, 235 with ADHD- Inattentive (ADHD-I) type, and 445 typical children[47]. The 3 GDS subtests (Delay, Vigilance, and Distractibility) yield 5 primary scores: Delay efficiency ratio (percentage of correct responses), Vigilance number of correct responses and number of commission errors (number of times the child pushed the response button when 1 was not followed by 9), and Distractibility and commission errors. In this study, typical children had a GDS composite standard score of 100, consistent with the normal mean of 100 in the 1983 standardization sample. Means for children with ADHD-C and ADHD-I were 70 and 78, respectively, approximately 2 standard deviations below the normal mean. There were statistically significant differences between Vigilance, Distractibility, and Delay mean composite standard scores and the 1983 standardization sample mean of 100 for children with ADHD-C and ADHD-I. No statistically significant differences in

Vigilance correct scores or Distractibility correct scores were reported between children with ADHD-C and ADHD-I. However, children with ADHD-C had more Distractibility commission errors ($p<0.01$), poorer Delay scores ($p<0.0001$), and Vigilance commission errors than children with ADHD-I ($p=0.01$). The authors concluded that children's ability to pay attention and inhibit impulsive behaviors, as measured by the GDS, has remained stable over the past 20 years.

The GDS studies identified in Hall et al.'s SLR reported mixed results on its clinical utility for aiding in the diagnosis of ADHD in children. The reported GDS sensitivities ranged from 49% to 90% with specificities ranging between 70% and 95% for ADHD diagnosis. The GDS Delay task was reported to aid in ADHD subgroup differentiation. The ADHD combined subgroup had greater impulsivity than inattentive subgroup but both groups had similar vigilance and distractibility impairments (Delay task or Delay + Distractibility tasks classification accuracies ~ 70%). The GDS classification accuracy increased to 72% when combined with the Wechsler Intelligence Scale for Children (WISC) Freedom from Distractibility (FD)/Working Memory Index and Processing Speed Index. However, predicting diagnosis for ADHD subgroups was low (PPV: 46%). Increases in classification accuracy for ADHD diagnosis were also reported when combining GDS and IQ scores (IQ-GDS discrepancy score). The classification accuracy for ADHD combined type in a study was 79% when using the GDS composite score alone with a < 90 cut off (PPV: 90%, NPV: 52%). Using the13-point IQ-GDS cut off increased the accuracy to 86% (sens: 90%, spf: 70%), the PPV to 91%, and the NPV to 67%. A similar study using this cut off identified 88% of children as having ADHD combined type. This sensitivity increased to 91% when the IQ-GDS was combined with a 11-point difference between IQ and the WISC FD index. GDS correctly identified 70% of the children without ADHD whereas WISC identified only 31%.

Garcia-Murillo et al.[8] reported results from a meta-analysis and a case-control study in 1 publication. The meta-analysis evaluated locomotor activity measures used in the diagnosis of ADHD. The studies identified evaluated the OPTAx (n=2 studies published in 1996 and 2012) and QbTest Plus (n=3 studies published in 2010, 2012, and 2014; QbTest's version for subjects 12 years and older). Both attention task performance recorders provide measurements of hyperactivity, impulsivity, and inattention to aid in the clinical assessment of ADHD. The studies evaluating OPTAx included 50 subjects (18 children) diagnosed with ADHD and 71 (11 children) without ADHD. Those evaluating QbTest Plus included 106 subjects (45 children) with ADHD and 71 (45 children) without ADHD. A statistically significant difference in locomotor activity was found between subjects with ADHD and controls (standard mean difference 0.92, 95% CI: 0.6, 1.20, $p<0.05$; heterogeneity: $I^2=21\%$, $p=0.28$). The case-control study used OPTAx to evaluate locomotor activities measures in children, adolescents, and adults[8]. The case-control study included a group of children and adolescents (62 cases with ADHD and 61 typically developing (TD) controls) and a group of adults (19 cases and 30 controls). ADHD groups differed statistically significantly from controls on all motion tracking parameters (number of head movements, displacement, head area, spatial complexity, and temporal scaling) except head immobility duration and head spatial complexity in children/adolescents. In both age groups, in the concurrent GNG task, RT variability was statistically significantly greater in cases than controls ($p<0.05$). In both age groups, Conners' scores were also statistically significantly higher in cases than in controls on all measures ($p<0.01$).

Garcia-Murillo et al.[8] also evaluated the convergent validity of the OPTAx system with the Conners' Parent/Teacher Rating Scales revised and Conners' Adult ADHD Rating Scale

(CAARS). Within children and adolescents, all GNG measures, except for response latency, correlated statistically significantly with all motion tracking measures, even after adjusting for age. Among adults, only RT variability measures correlated statistically significantly with motion tracking measures. In children and adolescents, poor correlations, albeit statistically significant, were observed between number of head movements, head displacement, and head area and Conners' parent DSM-hyperactive–impulsive subscale ($r=0.28$-0.31, $p<0.01$). Head area was the only motion tracking measure correlated with the Conners' teacher DSM-hyperactive–impulsive subscale ($r=0.30$, $p<0.01$; poorly correlated). The correlations between head spatial complexity and head temporal scaling with Conners' parent or teacher rating scales were poor and not statistically significant. Among adults, the correlations between number of head movements, head displacement, and head area with the CAARS DSM-hyperactive–impulsive subscale were poor to moderate ($r=0.45$-0.57) and statistically significant ($p<0.01$).

Hurford et al.[9] conducted an RCT in 122 children to examine the diurnal assumptions (test-retest reliability) and concurrent validity of the T.O.V.A. in elementary students. The T.O.V.A. is recommended to be taken before 1 pm to match the time of data collection for the normative data. It produces 5 scores: omission and commission errors, RT and RT variability, and an overall ADHD score (D-prime (DP) score, used to differentiate persons with and without ADHD). Children were tested in 4 groups based on time of day (morning or afternoon) for the 1st and 2nd T.O.V.A. administration. The test-retest reliability from 1st to 2nd administration was consistent across groups, and there were no statistically significant differences between groups. Across groups, Pearson's correlation coefficients showed moderate reliability ($r=0.64$-0.74) for the total score of T.O.V.A. variables across groups. The authors also compared the 4 groups on the consistency of the T.O.V.A. interpretation (TI) and its ADHD score included at the 1st to 2nd administration. The TI provides a comparison of the subject's performance with a normative sample by age and gender with individuals who do not have an attention problem and the ADHD score is a comparison of the subject's performance with individuals from the normative sample who have been independently diagnosed with ADHD[52].The results showed at least 73% of consistency between the 2 administrations for the TI and ADHD decisions regardless of group. The Vanderbilt ADHD Diagnostic Teacher Rating Scale was used to evaluate concurrent validity. Children who were identified as ADHD with the Vanderbilt (n=28) were consistently classified as ADHD (89.3%, 25/28 (sensitivity)) on the T.O.V.A. regardless of time of day. Hall et al.'s SLR identified a study were T.O.V.A. was shown to complement the clinical assessment process. The study evaluated T.O.V.A. and the Revised Conners' Teacher RS (CTRS-R) and showed that these assessments did not identify identical groups of children, with each correctly classifying children misclassified by the other assessment. Any one T.O.V.A. measure > 1.5 SD correctly classified 80% of the ADHD group and 72% of the non-ADHD group. The authors concluded that the published studies identified "support that [attention task performance recorders] provide an objective method to assess attention and impulsivity, but there are mixed reports on whether they are a useful adjunct to clinical practice."

Reh et al.[35] conducted a cross-sectional study in 930 German children to analyze the QbTest's factor structure to determine if it captures the 3 core ADHD symptoms (hyperactivity, inattention, and impulsiveness), evaluate its internal reliability, and construct validity (convergent and discriminant) with the Conners' 3rd Parent/Teacher Rating Scales. QbTest's exploratory factors analysis and internal reliability evaluation were conducted in a subset of 828 German children referred for ADHD assessment. It identified 3 factors that explained 76% of the total variance:

hyperactivity, accounted for 49.13% of the total variance with 5 QbTest variables conceptually related to motor activity/motion (i.e., time active, distance, area, microevents, motion simplicity), inattention, accounted 14.43% of the variance, with 3 variables conceptually related to inattention (i.e., omission errors, RT, RT variability), and impulsiveness accounted for 12.11% of the total variance with variables conceptually related to impulsivity (i.e., commission errors, multiresponse, anticipatory). Cronbach's alphas, which measure internal reliability, were excellent for hyperactivity (α =0.95) and inattention (α =0.76) but below of the recommended cutoff for impulsivity (α =0.60)[53]. Across all 3 subscales, QbTest scores decreased with age and gender was found to statistically significantly influence ADHD symptom severity. The authors used the multitrait–multimethod (MTMM) matrix approach in the rest of the children, a subset of 102 "strictly" diagnosed with ADHD, to examine convergent validity and discriminant/divergent. Hyperactivity was the only QbTest factor that had a statistically significant positive correlation with teacher ratings of hyperactive behavior ($r = 0.27$, $p < 0.01$) on the Conners' DSM hyperactivity/impulsivity subscale. None of the QbTest factors were statistically significantly correlated with the Conners' parent ratings of inattentive or hyperactive/impulsive behavior. Regarding discriminant validity, there were no statistically significant correlations found between QbTest factors and Conners' parent and teacher ratings of Peer Relations, the different construct selected for this evaluation. Hirsch and Christiansen[34] did a similar evaluation of the QbTest Plus using a sample of 773 German subjects (age range: 12-76) for confirmatory and exploratory factor analyses and a second sample of 297 subjects (age range: 16-60) to examine concurrent and discriminant validity using MTMM analysis. The authors concluded that the criterion for convergent validity was fulfilled and discriminant validity was partially supported. They also reported that omission errors and RT variability (part of the inattention factor) were able to discriminate between subjects with and without ADHD. The internal reliability of the QbTest Plus was also verified. In this sample, the classification accuracy was 76.4% (sens: 90%, spf: 45%, PPV: 79%, NPV: 67%, prevalence of 69.4% (206/297)).

In another study in adults, adding QbTest to clinical rating scales improved the differentiation of ADHD and autism spectrum disorders (ASD) in English adults[33]. In this study, the combination of CAARS-E and Autism Quotient 10-item version successfully classified 81% of participants (16% of the ADHD sample and 24% of the ADS sample were misclassified). Adding the QbTest scores improved the classification accuracy to 90% ($p < 0.05$) and only 2 individuals with ADHD (6%) and 4 with ASD (16%) were incorrectly classified. The authors identified a Q-score of 1.12 associated with 84% sensitivity and 80% specificity to ADHD and suggested that this may be a useful cut-off predictive of ADHD when the sample comprises individuals with ADHD and ASD diagnoses.

A validity study among adults from Germany, The Netherlands, and Sweden (97 cases and 112 controls) reported group differences in QbTest raw scores where in all parameters, except for commission errors, the ADHD group scored worse compared with the control group[32]. The QbTest hyperactivity and inattention factors contributed statistically significantly in differentiating the ADHD and non-ADHD groups. The QbTest correctly classified 70.3% of the subjects (sens: 56%, spf: 83%, PPV: 74%, and NPV: 68%) and increased to 91% (sens: 91%, spf: 91%, PPV: 92%, and NPV: 90%) when combined with self-reports of ADHD symptoms severity.

Hall et al.[39] conducted a retrospective cohort study where they examined the records of 40 patients with ADHD diagnosed without the QbTest and 40 diagnosed with the QbTest. They showed that

statistically significantly fewer clinician consultations (mean 2.18 vs. 3.05; $p<0.02$) were required to confirm the diagnosis of ADHD when the QbTest was used to aid in the assessment in comparison to standard assessment as usual.

Ramtvedt and Sundet[21] evaluated construct validity of the QbTest with the Conners' teacher rating scale in a sample of 36 Norwegian children with ADHD. All of the correlations of the QbTest inattention factor or QbTest activity factor with the Conners' teacher ratings of inattentive or hyperactive/impulsive behavior were poor (r=0.10-0.42). In a double-blind RCT evaluating atomoxetine (ATX) in 125 German children with ADHD, correlations between the QbTest factors and ADHD-RS were poor at baseline[19]. For changes from baseline, the highest correlations reported were moderate and were between overall accuracy (error rate) and ADHD-RS hyperactivity-impulsivity subscore, inattention subscore, and total score (r=0.56-0.61).

Bijlenga et al.[38] compared QbTest and ADHD-RS results from a pretest-posttest observational study evaluating stimulant treatment in 145 Dutch adults with ADHD. Self-reported ADHD symptoms were poorly correlated to QbTest scores, with the total scores of the ADHD-RS and QbTest yielding the highest correlation (r=0.33, $p< .01$). The QbTest factors were poorly correlated with each other (r=-0.1-0.24) while the ADHD-RS subscales had poor to moderate correlations to each other (r=0.41-0.63, $p\leq0.01$). When patients were classified in 3 categories (clinical improvement, no change, or deterioration), the ADHD-RS and QbTest agreed in 47% of the cases (36/77). There were 32% (25/77) of the patients who showed improvement in the QbTest but did not show in the ADHD-RS. QbTest objectified an improvement in 54% (25/46) of patients who subjectively did not report an improvement on the ADHD-RS.

Vogt and Williams[24] evaluated the internal reliability of the QbTest in a pretest-posttest trial. There were good to excellent statistically significant correlations between the 4 activity measures (time active, distance, area, and microevents; r=0.83–0.96, $p<0.001$). Omission errors (attention) had moderate and statistically significant correlations with the 4 activity measures (r=0.53–0.57, $p<0.001$) and with commission errors (impulsivity; r=0.51, $p<0.001$). Activity measures were poorly correlated with normalized RT variation and commission errors.

Ulberstad et al.[31] evaluated the test-retest reliability and concurrent validity of the QbCheck in a sample of 25 adolescents and adults (11 with ADHD and 14 without ADHD) from Germany. The test-retest reliability assessment showed excellent to good ICC for all 5 variables: microevents ($ICC_{2,1}$=0.90), omission errors ($ICC_{2,1}$=0.84), commission errors ($ICC_{2,1}$=0.82), RT ($ICC_{2,1}$=0.96), and RT variability ($ICC_{2,1}$=0.88). The concurrent validity of the QbCheck was excellent. The correlation between the microevents variables measured with the cameras used for QbCheck and QbTest was excellent and statistically significant ($r =0.91$ $p<0.001$). The authors also evaluated convergent validity and diagnostic validity of the QbCheck in a second sample of 140 adolescents and adults (67 with ADHD and 73 without ADHD) from Germany, Sweden, and Alabama. Data to assess convergent validity was only available from subjects with ADHD (i.e., those with results from QbCheck and QbTest). The correlations between the QbCheck and corresponding QbTest variables were moderate and statistically significant for all 5 variables (r=0.50-0.74). The sensitivity of the QbCheck was 82.6% and the specificity was 79.5%.

In summary, the evidence on the validity of the GDS was mostly summarized in Hall et al.'s SLR. The GDS scores had higher classification accuracies when used in combination with other rating

scales. The GDS was also evaluated in a large sample study of more than 1,200 children and showed consistent results with the standardization sample obtained more than 35 years ago. For attention task performance recorders that can measure activity, the meta-analysis showed that subjects with ADHD had higher activity measurements compared to those without ADHD in the OPTAx and QbTest. Similar results were observed in the case-control study evaluating the OPTAx. Although the evidence for T.O.V.A. is limited, T.O.V.A. seem to have a moderate test-retest reliability and a good concurrent validity when compared to the Vanderbilt RS, although the sample size of the study was small. Only 1 study evaluated QbCheck and compared it to QbTest. It was shown to have excellent to good test-retest reliability, excellent concurrent validity, moderate convergent validity, and a sensitivity of ~83% and specificity of ~80%.

There was more evidence available on QbTest as an aid in the clinical assessment of ADHD. The results show that this attention task performance recorder had good convergent and discriminant validity. In general, correlations of the QbTest with Conners' parent, teacher rating scales, or the ADHD-RS were found to be poor. However, the classification accuracy of the QbTest was reported to be up to 90% with higher values when the QbTest scores were combined with rating scales in the clinical assessment. The QbTest was reported to also have a good internal reliability. Overall, these studies show favorable results. All the evidence evaluating QbTest as an aid in the clinical assessment of ADHD was from outside the US, which limits the studies' generalizability to the US population given that the practice of medicine is different as well as the population awareness and acceptability of ADHD. In addition, the reference sample used to classify ADHD was drawn from Sweden. More confirmatory studies are needed evaluating these devices in the US to assess its clinical utility as an aid in the clinical assessment of ADHD.

Aid in the Evaluation of Treatment Interventions for ADHD

One (1) SLR[6] and 17 publications (reporting on 15 studies) evaluating attention task performance recorders as part of the assessment of treatment interventions for patients with ADHD were identified. Out of the 15 studies, 11 were on pharmacological interventions[10-12,14-16,18-21,24,38,40]. T.O.V.A. was used in 5 studies (4 RCTs[10-13] and a single-arm trial[25]) and QbTest in 10 studies (6 RCTs[14-21], 2 clinical trials[24,26], and 2 observational studies[38,40]). These are the only attention task performance recorders with clearance for this indication to aid in the evaluation of treatment interventions for ADHD.

Hall et al.[6]'s SLR evaluated the evidence for attention task performance recorders that have been used for aiding in medication monitoring for children and young people with ADHD. They identified 9 publications using GDS (n=2), T.O.V.A. (n=4), QbTest (n=1), and OPTAx System (also known as McLean Motion and Attention Test, n=2). They reported that T.O.V.A. showed limited sensitivity to medication effects, with only impulsivity scores improving as a result of the medication. QbTest and OPTAx were shown to be clinically useful in supporting titration. The authors concluded that "there was a strong evidence base for the use of objective measures of activity to aid ADHD/non-ADHD group differentiation, which appears sensitive to medication effects".

Two (2) double-blind crossover RCTs evaluating extended-release methylphenidate (MPH) in children with ADHD reported statistically significant improvement in T.O.V.A. RT, RT variability, and ADHD scores in those treated with MPH compared to placebo[10,11]. Williamson et al[12]. pooled the data from these 2 RCTs to determine if there were any differences in treatment

response in children with ADHD with and without learning disabilities. Children who received MPH had less RT variability than those who received placebo, regardless of the presence of learning disabilities. Results on RT and ADHD scores were not reported. A double-blind RCT explored the potential use of binaural auditory beat stimulation to reduce inattention in 20 children and adolescents with ADHD concurrently taking stimulant medication[13]. There were no significant group differences on the T.O.V.A. omission errors scores post-intervention. However, this RCT had only 19% statistical power to detect group differences. Ezra et al[25]. conducted a single-arm trial in 15 subjects who developed an attention deficit disorder and slowing of RT at the time of exposure to mold toxins. A statistically significantly improvement in attention span, RT, RT variability, and an overall D prime score after receiving 10 treatment sessions of mild hyperbaric oxygen treatment.

QbTest was used in the evaluation of effectiveness of treatment interventions in reducing ADHD symptoms in 10 European studies. Most studies evaluated pharmacological interventions (5 RCTs[14-16,18,20], 1 clinical trial[24], and 2 observational studies[38,40]) and were conducted in children[14,18,20,24,26,40]. Statistically significant improvement in QbTest scores measuring the 3 core ADHD symptoms were observed in 5 trials and the observational study evaluating MPH[14,15,24], MPH/dextroamphetamine[20,38] (1 did not evaluate impulsivity[20]), and/or ATX[18]. In a double-blind RCT in 30 adults with ADHD, performance on the QbTest was not statistically significantly different between those treated with a cannabinoid medication and those using placebo[16]. There were 2 trials identified evaluating nonpharmacological interventions in adolescents and children. Anodal transcranial direct current stimulation (tDCS) of the prefrontal cortex treatment was studied in 15 adolescents with ADHD in a double-blind crossover RCT[17]. There was a statistically significant reduction in hyperactivity in the treated group compared with sham stimulation 7 days after the end of stimulation ($p=0.02$)[17]. Changes in QbTest scores did not reach statistical significance for measures of inattention or impulsivity in this trial. The effects of practicing target-shooting sport on the severity of hyperactivity and inattention symptoms were evaluated in a clinical trial of 128 Danish children using QbTest[26]. Statistically significant improvement was found only in the QbTest measurement of attention (RT variability and omission errors). This trial also observed a statistically significant improvement in the parent-rated ADHD-RS total score ($p=0.024$).

A Swedish study of 80 children with ADHD and 38 without it did not find the QbTest to have clinical utility in the diagnostic assessment of children with ADHD[40]. However, the same study also included 56 children with ADHD who were followed for 1 year. In this sample, the QbTest was found useful in predicting MPH treatment response 1 year later when parent SNAP-IV ratings were inconclusive. For children without clear results for inattention either using the SNAP-IV parent rating or the QbTest inattention, supplementing these scores with QbTest activity ratings resulted in a high sensitivity (98%, spf: 25%, PPV: 89%, NPV: 67%). Although the specificity was low, this combination resulted in only 1 false negative. As Tallberg et al. noted “false positive cases can be identified during the follow-up. On the other hand, having many false negatives regarded as non-responders would lead to many children not getting the chance to benefit from a medicine.”

The evidence on the ability of attention task performance recorders to aid in the evaluation of treatment interventions for ADHD is limited. The conclusion of the SLR evaluating the evidence in children support the use of attention task performance recorders that provide objective measures

of activity to "aid ADHD/non-ADHD group differentiation, which appears sensitive to medication effects"[6]. FDA's SLR included 5 of the 6 studies evaluating levels of activity used to capture effects of medications. QbTest was able to capture statistically significant improvement in QbTest scores measuring the 3 core ADHD symptoms in 6 studies evaluating traditional pharmacological interventions.

5.5 Overall Literature Review Conclusions

This SLR did not identify studies reporting adverse events related to the use of an attention task performance recorder itself. The evidence suggests that the challenge from these devices is the possibility of false negatives and false positives, which relates to their effectiveness, and could lead to harm if a misdiagnosis results in inappropriate or no treatment. The devices included in the LQD product code are heterogeneous in terms of their indications for use and design, preventing combining study results to draw meaningful conclusions regarding their effectiveness. Furthermore, there are attention task performance recorders that have clearance to measure RT without a diagnostic claim, be used as a cognitive screening tool, and provide objective measurements of hyperactivity, and/or attention and inhibitory control, which helps with the clinical assessment of ADHD. Some of these devices can also aid in the evaluation of treatment interventions in patients with ADHD. The evidence identified for most devices was sparse. Other limitations include studies' sample size and limited generalizability of results to the US population (~two fifths of the studies were conducted outside the US). In the US, awareness among health care providers, parents, and educators regarding ADHD symptoms has increased over time. In addition, the practice of medicine for the diagnosis and treatment of ADHD in the US may be different than in other countries. The "gold standard" for the validation of these devices are rating scales. Those used for children are completed by teachers and parents and provide information in different settings (school and home). Although they provide useful information, they rely on their interpretation and recall of the child behavior. Similar issues affect self-reports from adults. Translation and cultural interpretations of the questions in the rating scales can also affect the generalizability of the results on validity.

There are several implications with misdiagnosis of individuals with ADHD. These devices are used to provide objective measures of the core symptoms of ADHD, with some measuring hyperactivity and providing scores standardized by age. A false positive result can expose a child to unnecessary behavioral and pharmacological treatment, in addition to increased parental anxiety. In contrast, individuals with ADHD with a false negative test result will miss needed treatment (behavioral or pharmaceutical), interventions (or tools) at school (or work), understanding, and acceptance.

6. Risks to Health Identified through Medical Device Reports (MDRs)

6.1 Overview of the MDR System

The MDR system provides FDA with information on medical device performance from patients, health care professionals, consumers and mandatory reporters (manufacturers, importers and device user facilities). The FDA receives MDRs of suspected device-associated deaths, serious injuries, and certain malfunctions. The FDA uses MDRs to monitor device performance, detect potential device-related safety issues, and contribute to benefit-risk assessments of these products. MDRs can be used effectively to:

- Establish a qualitative snapshot of adverse events for a specific device or device type.
- Detect actual or potential device problems used in a “real world” setting/environment.

Although MDRs are a valuable source of information, this passive surveillance system has limitations, including the submission of incomplete, inaccurate, untimely, unverified, duplicated or biased data. In addition, the incidence or prevalence of an event cannot be determined from this reporting system alone due to potential under-reporting of events and lack of information about the frequency of device use. Finally, the existence of an adverse event report does not definitely establish a causal link between the device and the reported event. Because of these limitations, MDRs comprise only one of the FDA’s tools for assessing device performance. As such, MDR numbers and data should be taken in the context of the other available scientific information.

6.2 MDR Data: Attention Task Performance Recorder Devices (Product Code LQD)

Individual MDRs for attention task performance recorders are reported through FDA’s Manufacturer and User Facility Device Experience (MAUDE) database, which houses mandatory reports from medical device manufacturers, importers and user facilities, as well as voluntary reports from entities such as health care professionals, patients and consumers.

The FDA conducted queries of the MDR database on March 9, 2021, to identify adverse events related to the use of attention task performance recorders (product code LQD). The search was not timeframe restricted and included all MDRs reported under product code LQD. The search did not identify any relevant MDRs for attention task performance recorder devices.

7. Recall History

7.1 Overview of Recall Database

The Medical Device Recall database contains Medical Device Recalls classified since November 2002. Since January 2017, it may also include correction or removal actions initiated by a firm prior to review by the FDA. The status is updated if the FDA identifies a violation and classifies the action as a recall and again when the recall is terminated. FDA recall classification may occur after the firm recalling the medical device product conducts and communicates with its customers about the recall. Therefore, the recall information posting date ("create date") identified on the database indicates the date FDA classified the recall, it does not necessarily mean that the recall is new.

7.2 Recall Results: Attention Task Performance Recorder Devices

The FDA conducted queries of the Medical Device Recall database on March 11, 2021, to identify recalls related to attention task performance recorders (product code LQD). The search was not timeframe restricted and included all recalls reported under product code LQD. The search did not identify any relevant recalls for attention task performance recorder devices.

8. Summary

In light of the information available, the Panel will be asked to comment on whether attention task performance recorders under product code “LQD”

meet the statutory definition of a Class III device:

- insufficient information exists to determine that general and special controls are sufficient to provide reasonable assurance of its safety and effectiveness, and

- the device is life-supporting or life-sustaining, or for a use which is of substantial importance in preventing impairment of human health, or if the device presents a potential unreasonable risk of illness or injury;

or would be more appropriately regulated as Class II, in which:

- general and special controls, which may include performance standards, postmarket surveillance, patient registries and/or development of guidelines, are sufficient to provide reasonable assurance of safety and effectiveness;

or as Class I, in which:

- the device is subject only to general controls, which include registration and listing, good manufacturing practices (GMPs), prohibition against adulteration and misbranding, and labeling devices according to FDA regulations.

For the purposes of classification, FDA considers the following items, among other relevant factors, as outlined in 21 CFR 860.7(b):

1. The persons for whose use the device is represented or intended;

2. The conditions of use for the device, including conditions of use prescribed, recommended, or suggested in the labeling or advertising of the device, and other intended conditions of use;

3. The probable benefit to health from the use of the device weighed against any probable injury or illness from such use; and

4. The reliability of the device.

8.1 Special Controls

FDA believes that special controls, in addition to general controls, can be established to mitigate the risks to health identified, and provide a reasonable assurance of the safety and effectiveness of attention task performance recorders. Following are risk/mitigation tables, which outline the identified risks to health for this device type and the recommended controls to mitigate the identified risks, delineated by intended use:

Table 5: Summary of Risks to Health and Proposed Special Controls for Attention Task Performance Recorders Intended to Measure Reaction Time and Associated Patient Performance in Response to Attention Tasks, Without Aiding in Assessment or Diagnosis

Identified Risk	Recommended Mitigation Measure
Patient discomfort (e.g., visual or mental fatigue)	• Labeling
Incorrect or inaccurate measurements of reaction time or other attention tasks	• Non-clinical performance testing • Software verification, validation, and hazard analysis • Labeling

Table 6: Summary of Risks to Health and Proposed Special Controls for Attention Task Performance Recorders Intended to Aid in Assessment or Diagnosis of Specific Diseases or Conditions

Identified Risk	Recommended Mitigation Measure
Patient discomfort (e.g., visual or mental fatigue)	• Labeling
Incorrect or inaccurate results leading to inaccurate assessment or delayed diagnosis, both of which could result in inappropriate therapy or delay in treatment	• Clinical performance testing • Non-clinical performance testing • Software verification, validation, and hazard analysis • Labeling

The Panel will be asked whether this list is a complete and accurate list of the risks to health presented for attention task performance recorders and whether any other risks should be included in the overall risk assessment of the device type.

Based on the identified risks and recommended mitigation measures, FDA believes that the following special controls would provide reasonable assurance of safety and effectiveness for attention task performance recorders *intended to measure reaction time and associated patient performance in response to attention tasks only without aiding in assessment or diagnosis*:

1. The technical parameters of the device's hardware and software must be fully characterized and be accompanied by appropriate non-clinical testing:
    a. Hardware specifications must be provided. Appropriate verification, validation and hazard analysis must be performed, including applicable electrical safety testing.
    b. Software, including any proprietary algorithm(s) used by the device to measure reaction time and output other measures of attention, associated activities and related task performance, must be described in detail in the Software Requirements Specification (SRS) and Software Design Specification (SDS). Appropriate software verification, validation and hazard analysis must be performed.

2. Non-clinical device performance evaluation must demonstrate accurate and precise measurement of patient reaction times in response to task stimuli.

3. The labeling must include:
    a. A warning that the device is not intended to aid in patient assessment or diagnosis of specific diseases or conditions.

b. Any instructions technicians must convey to patients regarding safe and effective administration of the specific tasks and collection of task performance data.

Based on the identified risks and recommended mitigation measures, FDA believes that the following special controls would provide reasonable assurance of safety and effectiveness for attention task performance recorders *intended to measure reaction time and associated patient performance in response to attention tasks for the aid in assessment or diagnosis of specific diseases or conditions*:

1. Clinical device performance evaluation must validate that the device outputs accurately and precisely assess patient symptomology associated with the specific disease or condition for which the device is intended to assess or diagnose. The testing must:
    a. Evaluate agreement between device output and patient symptomology.
    b. Evaluate device test-retest reliability.
    c. Describe construction of any normative or reference database, which includes the following:
        i. How the clinical work-up was completed to define the reference population, including the establishment of inclusion and exclusion criteria.
        ii. Statistical methods and model assumptions used.

2. The technical parameters of the device's hardware and software must be fully characterized and be accompanied by appropriate non-clinical testing:
    a. Hardware specifications must be provided. Appropriate verification, validation and hazard analysis must be performed, including applicable electrical safety testing.
    b. Software, including any proprietary algorithm(s) used by the device to measure reaction time and output other measures of attention, associated activities and related task performance, must be described in detail in the Software Requirements Specification (SRS) and Software Design Specification (SDS). Appropriate software verification, validation and hazard analysis must be performed.

3. Non-clinical device performance evaluation must demonstrate accurate and precise measurement of patient reaction times in response to task stimuli.

4. The labeling must include:
    a. A summary of any clinical testing conducted to demonstrate that the device outputs accurately and precisely assess patient symptomology associated with the specific disease or condition for which the device is intended to assess or diagnose. The summary of testing must include the following:
        i. Agreement between device output and patient symptomology.
        ii. Device test-retest reliability.
        iii. A description of any normative or reference database, which includes the following:
            1. How the clinical work-up was completed to define the reference population, including the establishment of inclusion and exclusion criteria.
            2. How reference values will be reported to the user.

3. Representative screenshots and reports that will be generated to provide the user results and reference data.
4. Statistical methods and model assumptions used.
5. Whether or not the database was adjusted due to differences in age, gender, or other factors.

b. A warning that the device is intended to aid in patient assessment or diagnosis by a trained physician and is not intended for stand-alone use.
c. Any instructions that technicians must convey to patients regarding safe and effective administration of the specific tasks and collection of task performance data.

If the Panel believes that Class II is appropriate for attention task performance recorders under product code "LQD," the panel will be asked whether the identified special controls appropriately mitigate the identified risks to health and whether additional or different special controls are recommended.

8.2 Overview of Proposed Classification/FDA Recommendation

Based on the safety and effectiveness information gathered by the FDA, the identified risks to health and recommended mitigation measures, we recommend that attention task performance recorders indicated for use for measuring reaction time in response to attention tasks be regulated as Class II devices.

882.1490 Attention task performance recorder.

(a) *Identification.*

An attention task performance recorder is a device intended to measure reaction time and associated patient performance in response to attention tasks. The device may or may not be used to aid in the assessment of specific clinical conditions.

(b) *Classification.*

(1) Class II (special controls), when intended to measure reaction time and associated patient performance in response to attention tasks only without aiding in assessment or diagnosis:

1. The technical parameters of the device's hardware and software must be fully characterized and be accompanied by appropriate non-clinical testing:
    a. Hardware specifications must be provided. Appropriate verification, validation and hazard analysis must be performed, including applicable electrical safety testing.
    b. Software, including any proprietary algorithm(s) used by the device to measure reaction time and output other measures of attention, associated activities and related task performance, must be described in detail in the Software Requirements Specification (SRS) and Software Design Specification (SDS). Appropriate software verification, validation and hazard analysis must be performed.

2. Non-clinical device performance evaluation must demonstrate accurate and precise measurement of patient reaction times in response to task stimuli.

3. The labeling must include:
    a. A warning that the device is not intended to aid in patient assessment or diagnosis of specific diseases or conditions.
    b. Any instructions technicians must convey to patients regarding safe and effective administration of the specific tasks and collection of task performance data.

(2) Class II (special controls), when intended to measure reaction time and associated patient performance in response to attention tasks for the aid in assessment or diagnosis of specific diseases or conditions:

1. Clinical device performance evaluation must validate that the device outputs accurately and precisely assess patient symptomology associated with the specific disease or condition for which the device is intended to assess or diagnose. The testing must:
    a. Evaluate agreement between device output and patient symptomology.
    b. Evaluate device test-retest reliability.
    c. Describe construction of any normative or reference database, which includes the following:
        i. How the clinical work-up was completed to define the reference population, including the establishment of inclusion and exclusion criteria.
        ii. Statistical methods and model assumptions used.

2. The technical parameters of the device's hardware and software must be fully characterized and be accompanied by appropriate non-clinical testing:
    a. Hardware specifications must be provided. Appropriate verification, validation and hazard analysis must be performed, including applicable electrical safety testing.
    b. Software, including any proprietary algorithm(s) used by the device to measure reaction time and output other measures of attention, associated activities and related task performance, must be described in detail in the Software Requirements Specification (SRS) and Software Design Specification (SDS). Appropriate software verification, validation and hazard analysis must be performed.

3. Non-clinical device performance evaluation must demonstrate accurate and precise measurement of patient reaction times in response to task stimuli.

4. The labeling must include:
    a. A summary of any clinical testing conducted to demonstrate that the device outputs accurately and precisely assess patient symptomology associated with the specific disease or condition for which the device is intended to assess or diagnose. The summary of testing must include the following:
        i. Agreement between device output and patient symptomology.

ii. Device test-retest reliability.
iii. A description of any normative or reference database, which includes the following:
    1. How the clinical work-up was completed to define the reference population, including the establishment of inclusion and exclusion criteria.
    2. How reference values will be reported to the user.
    3. Representative screenshots and reports that will be generated to provide the user results and reference data.
    4. Statistical methods and model assumptions used.
    5. Whether or not the database was adjusted due to differences in age, gender, or other factors.

b. A warning that the device is intended to aid in patient assessment or diagnosis by a trained physician and is not intended for stand-alone use.
c. Any instructions technicians must convey to patients regarding safe and effective administration of the specific tasks and collection of task performance data.

Based on the available scientific evidence, the FDA will ask the Panel for their recommendation on the appropriate classification of the attention task performance recorders under product code "LQD."

Appendix A: Literature Search Results for Attention Task Performance Recorders

A systematic literature review was conducted in an effort to gather any published information regarding the safety and effectiveness of attention task performance recorders under product code "LQD" from January 1, 2010, and through December 31, 2020. We searched 2 electronic databases (MEDLINE and Embase) using search terms limited to currently cleared attention task performance recorders and keywords from their indications for use (IFU); The searches were limited to studies published in English, with abstracts, and conducted in humans. The following search terms were used:

(1) PubMed: (Hyperactivity OR impulsivity OR inattention OR inhibitory control OR memory task OR reaction time OR working memory OR ((choice OR cognitive) AND discrimination) OR perceptual speed OR episodic memory OR crystallized abilities OR ADHD OR Attention Deficit Hyperactivity Disorders OR "Attention Deficit Disorder with Hyperactivity"[Mesh] OR vision OR intelligence) AND ((TOVA OR t.o.v.a. OR Test Of Variables Of Attention [tw]) OR ((cognitive OR neuro) AND (DANA [tw] OR Defense Automated Neurobehavioral Assessment [tw])) OR QBcheck OR QBtest OR "Qb test"[tw] OR "QB test"[tw] OR "qb test"[tw] OR "optax"[All Fields] OR Dynavision OR ((“Fagan"[Title/Abstract] NOT nomogram)) OR ((cognitive OR neuro OR attention) AND "Gordon"[Title/Abstract]) OR MMAT OR "McLean Motion and attention test" [all fields] OR "Quotient ADHD system" [all fields] OR "Quantified Behavior Test"[All Fields] OR "Quantified Behavioral Test" [All Fields]) AND "2010/01/01"[PDat] : "2020/12/31"[PDat] AND English[lang] AND hasabstract[text] AND "humans"[MeSH Terms]

(2) PubMed search terms for references not indexed at the time of the search: (Hyperactivity OR impulsivity OR inattention OR inhibitory control OR memory task OR reaction time OR working memory OR ((choice OR cognitive) AND discrimination) OR perceptual speed OR episodic memory OR crystallized abilities OR ADHD OR Attention Deficit Hyperactivity Disorders OR "Attention Deficit Disorder with Hyperactivity"[Mesh] OR vision OR intelligence) AND ((TOVA OR t.o.v.a. OR Test Of Variables Of Attention [tw]) OR ((cognitive OR neuro) AND (DANA [tw] OR Defense Automated Neurobehavioral Assessment [tw])) OR QBcheck OR QBtest OR "Qb test"[tw] OR "QB test"[tw] OR "qb test"[tw] OR "optax"[All Fields] OR Dynavision OR (("Fagan"[Title/Abstract] NOT nomogram)) OR ((cognitive OR neuro OR attention) AND "Gordon"[Title/Abstract]) OR MMAT OR "McLean Motion and attention test" [all fields] OR "Quotient ADHD system" [all fields] OR "Quantified Behavior Test"[All Fields] OR "Quantified Behavioral Test" [All Fields]) AND "2010/01/01"[PDat]: "2020/12/31"[PDat] AND English[lang] AND hasabstract[text] NOT Medline[SB]

(3) EMBASE: ('hyperactivity'/exp OR 'hyperaction' OR 'hyperactivity' OR 'motor hyperactivity' OR 'impulsiveness'/exp OR 'behavior, impulsive' OR 'behaviour, impulsive' OR 'impulsive behavior' OR 'impulsive behaviour' OR 'impulsiveness' OR 'impulsivity' OR 'inattention'/exp OR 'inhibitory control'/exp OR 'memory task' OR 'reaction time'/exp OR 'overall reaction time' OR 'reaction latency' OR 'reaction time' OR 'response latency' OR 'response time' OR 'stimulus-response time' OR 'total reaction time' OR 'working memory'/exp OR 'memory, working' OR 'working memory' OR 'cognitive discrimination' OR 'choice discrimination' OR 'perceptual speed'/exp OR 'episodic memory'/exp OR 'episodic encoding' OR 'episodic memories' OR 'episodic memory' OR 'event memories' OR 'event memory' OR 'memory, episodic' OR 'crystallized intelligence'/exp OR 'crystallized abilities' OR 'attention deficit disorder'/exp OR 'adhd' OR 'attention deficit' OR 'attention deficit and disruptive behavior disorders' OR 'attention deficit and disruptive behaviour disorders' OR 'attention deficit disorder' OR 'attention deficit disorder with hyperactivity' OR 'attention deficit hyperactivity disorder' OR 'vision'/exp OR 'capacity, visual' OR 'central vision' OR 'figural aftereffect' OR 'half vision' OR 'ocular vision' OR 'optic perception' OR 'perception, optic' OR 'perception, visual' OR 'perceptual closure' OR 'phosphene' OR 'phosphenes' OR 'twilight vision' OR 'vision' OR 'vision, entoptic' OR 'vision, ocular' OR 'visual capacity' OR 'visual detection' OR 'visual function' OR 'visual perception' OR 'visual performance' OR 'visual process' OR 'visual sensation') AND ('tova':dn OR 'tova'/dn OR 'tova':ti OR 'tova':ab OR 't.o.v.a.':dn OR 't.o.v.a.':ti OR 't.o.v.a.':ab OR 'test of variables of attention'/exp OR 'tova score' OR 'tova test' OR 'test of variables of attention' OR 'DANA':dn OR 'DANA'/dn OR 'DANA':ti OR 'DANA':ab OR ‘Defense Automated Neurobehavioral Assessment’ OR (defense AND automated AND neurobehavioral AND assessment) OR 'QBcheck':dn OR 'QBcheck’/dn OR 'QBcheck’:ti OR 'QBcheck’:ab OR 'QBtest’:dn OR 'QBtest’/dn OR 'QBtest’:ti OR 'QBtest’:ab OR 'optax':dn OR 'optax'/dn OR 'optax':ti OR 'optax':ab OR 'Dynavision':dn OR 'Dynavision'/dn OR 'Dynavision':ti OR 'Dynavision':ab OR 'Fagan':dn OR 'Fagan'/dn OR 'Fagan':ti OR 'Fagan':ab OR 'Gordon':dn OR 'Gordon'/dn OR 'Gordon':ti OR 'Gordon':ab OR ('MMAT':dn OR 'MMAT'/dn OR 'MMAT':ti OR 'MMAT':ab OR 'McLean Motion and attention test':dn OR 'McLean Motion and attention test'/dn OR 'McLean Motion and attention

test':ti OR 'McLean Motion and attention test':ab OR 'Quotient ADHD system':dn OR 'Quotient ADHD system'/dn OR 'Quotient ADHD system':ti OR 'Quotient ADHD system':ab OR 'quantified behavior test' OR 'quantified behavior test':dn OR 'quantified behavior test'/dn OR 'quantified behavior test':ti OR 'quantified behavior test':ab OR 'quantified behavioral test' OR 'quantified behavioral test':dn OR 'quantified behavioral test'/dn OR 'quantified behavioral test':ti OR 'quantified behavioral test':ab)) AND [2010-2020]/py AND [english]/lim AND [abstracts]/lim AND [humans]/lim AND ('article'/it OR 'article in press'/it OR 'review'/it)

Appendix B: Flow Diagram of Systematic Literature Review Search Results

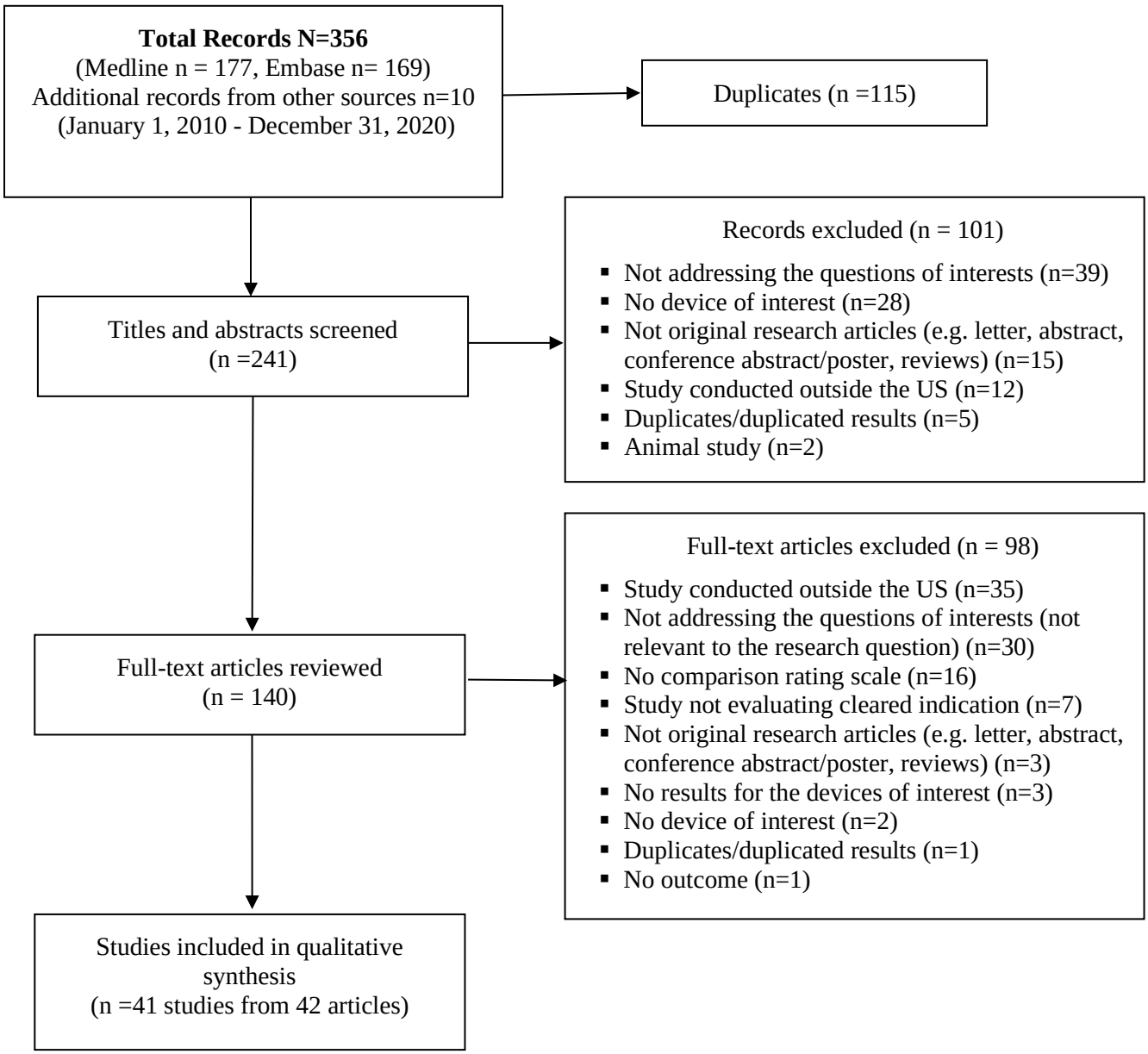


Figure 3. Flow Diagram of Article Retrieval and Selection

References

1. Danielson ML, Bitsko RH, Ghandour RM, Holbrook JR, Kogan MD, Blumberg SJ. Prevalence of Parent-Reported ADHD Diagnosis and Associated Treatment Among U.S. Children and Adolescents, 2016. *J Clin Child Adolesc Psychol.* 2018;47(2):199-212.
2. American Psychiatric Association, Subcommittee on Children and Adolescents with Attention-Deficit/Hyperactivity Disorder APASoCaAwA-DH. *Diagnostic and Statistical Manual of Mental Disorders.* 5th ed. Arlington, VA: American Psychiatric Association; 2013.
3. Danielson ML, Visser SN, Chronis-Tuscano A, DuPaul GJ. A National Description of Treatment among United States Children and Adolescents with Attention-Deficit/Hyperactivity Disorder. *J Pediatr.* 2018;192:240-246.e241.
4. Wolraich ML, Hagan JF, Jr., Allan C, et al. Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. *Pediatrics.* 2019;144(4).
5. Visser SN, Danielson ML, Wolraich ML, et al. Vital Signs: National and State-Specific Patterns of Attention Deficit/Hyperactivity Disorder Treatment Among Insured Children Aged 2-5 Years - United States, 2008-2014. *MMWR Morb Mortal Wkly Rep.* 2016;65(17):443-450.
6. Hall CL, Valentine AZ, Groom MJ, et al. The clinical utility of the continuous performance test and objective measures of activity for diagnosing and monitoring ADHD in children: a systematic review. *European Child and Adolescent Psychiatry.* 2016;25(7):677-699.
7. Morgan C, Honan I, Allsop A, Novak I, Badawi N. Psychometric Properties of Assessments of Cognition in Infants With Cerebral Palsy or Motor Impairment: A Systematic Review. *Journal of pediatric psychology.* 2019;44(2):238-252.
8. García Murillo L, Cortese S, Anderson D, Di Martino A, Castellanos FX. Locomotor activity measures in the diagnosis of attention deficit hyperactivity disorder: Meta-analyses and new findings. *J Neurosci Methods.* 2015;252:14-26.
9. Hurford DP, Lasater KA, Erickson SE, Kiesling NE. Examination of the Diurnal Assumptions of the Test of Variables of Attention for Elementary Students. *Journal of Attention Disorders.* 2013;17(3):208-214.
10. Wigal SB, Wigal T, Schuck S, et al. Academic, behavioral, and cognitive effects of OROS® methylphenidate on older children with attention-deficit/hyperactivity disorder. *Journal of Child and Adolescent Psychopharmacology.* 2011;21(2):121-131.
11. Murray DW, Childress A, Giblin J, Williamson D, Armstrong R, Starr HL. Effects of OROS methylphenidate on academic, behavioral, and cognitive tasks in children 9 to 12 years of age with attention-deficit/hyperactivity disorder. *Clinical Pediatrics.* 2011;50(4):308-320.
12. Williamson D, Murray DW, Damaraju CV, Ascher S, Starr HL. Methylphenidate in Children With ADHD With or Without Learning Disability. 2014;18((Williamson D.; Starr H.L., HStarr@its.jnj.com) Janssen Scientific Affairs, LLC, Raritan, NJ, United States):95-104.

13. Kennel S, Taylor AG, Lyon D, Bourguignon C. Pilot Feasibility Study of Binaural Auditory Beats for Reducing Symptoms of Inattention in Children and Adolescents with Attention-Deficit/Hyperactivity Disorder. *Journal of Pediatric Nursing.* 2010;25(1):3-11.

14. Gunther T, Kahraman-Lanzerath B, Knospe EL, Herpertz-Dahlmann B, Konrad K. Modulation of attention-deficit/hyperactivity disorder symptoms by short- and long-acting methylphenidate over the course of a day. *J Child Adolesc Psychopharmacol.* 2012;22(2):131-138.

15. Ginsberg Y, Hirvikoski T, Grann M, Lindefors N. Long-term functional outcome in adult prison inmates with ADHD receiving OROS-methylphenidate. *Eur Arch Psychiatry Clin Neurosci.* 2012;262(8):705-724.

16. Cooper RE, Williams E, Seegobin S, Tye C, Kuntsi J, Asherson P. Cannabinoids in attention-deficit/hyperactivity disorder: A randomised-controlled trial. *European Neuropsychopharmacology.* 2017;27(8):795-808.

17. Soff C, Sotnikova A, Christiansen H, Becker K, Siniatchkin M. Transcranial direct current stimulation improves clinical symptoms in adolescents with attention deficit hyperactivity disorder. *Journal of Neural Transmission.* 2017;124(1):133-144.

18. Wehmeier PM, Schacht A, Wolff C, Otto WR, Dittmann RW, Banaschewski T. Neuropsychological outcomes across the day in children with attention-deficit/hyperactivity disorder treated with atomoxetine: results from a placebo-controlled study using a computer-based continuous performance test combined with an infra-red motion-tracking device. *J Child Adolesc Psychopharmacol.* 2011;21(5):433-444.

19. Wehmeier PM, Schacht A, Ulberstad F, et al. Does atomoxetine improve executive function, inhibitory control, and hyperactivity? Results from a placebo-controlled trial using quantitative measurement technology. *J Clin Psychopharmacol.* 2012;32(5):653-660.

20. Ramtvedt BE, Roinas E, Aabech HS, Sundet KS. Clinical gains from including both dextroamphetamine and methylphenidate in stimulant trials. *J Child Adolesc Psychopharmacol.* 2013;23(9):597-604.

21. Ramtvedt BE, Sundet K. Relationships between computer-based testing and behavioral ratings in the assessment of attention and activity in a pediatric ADHD stimulant crossover trial. *Clin Neuropsychol.* 2014;28(7):1146-1161.

22. LaValle CR, Carr WS, Egnoto MJ, et al. Neurocognitive Performance Deficits Related to Immediate and Acute Blast Overpressure Exposure. *Frontiers in Neurology.* 2019;10.

23. Roach EB, Bleiberg J, Lathan CE, Wolpert L, Tsao JW, Roach RC. AltitudeOmics: Decreased reaction time after high altitude cognitive testing is a sensitive metric of hypoxic impairment. *NeuroReport.* 2014;25(11):814-818.

24. Vogt C, Williams T. Early identification of stimulant treatment responders, partial responders and non-responders using objective measures in children and adolescents with hyperkinetic disorder. *Child and Adolescent Mental Health.* 2011;16(3):144-149.

25. Ezra N, Dang K, Heuser G. Improvement of attention span and reaction time with hyperbaric oxygen treatment in patients with toxic injury due to mold exposure. *European Journal of Clinical Microbiology and Infectious Diseases.* 2011;30(1):1-6.

26. Gohr Månsson A, Elmose M, Mejldal A, Dalsgaard S, Roessler KK. The effects of practicing target-shooting sport on the severity of inattentive, hyperactive, and impulsive symptoms in children: a non-randomised controlled open-label study in Denmark. *Nordic Journal of Psychiatry.* 2019;73(4-5):233-243.

27. Picha K, Quintana C, Glueck A, Hoch M, Heebner NR, Abt JP. Reliability of 5 Novel Reaction Time and Cognitive Load Protocols. *Journal of sport rehabilitation.* 2018;27(5):1-4.

28. Wells AJ, Hoffman JR, Beyer KS, et al. Reliability of the dynavision™ d2 for assessing reaction time performance. *J Sports Sci Med.* 2014;13(1):145-150.

29. Haran FJ, Dretsch MN, Bleiberg J. Performance on the Defense Automated Neurobehavioral Assessment across controlled environmental conditions. *Applied neuropsychology Adult.* 2016;23(6):411-417.

30. Lathan C, Spira JL, Bleiberg J, Vice J, Tsao JW. Defense Automated Neurobehavioral Assessment (DANA)-psychometric properties of a new field-deployable neurocognitive assessment tool. *Mil Med.* 2013;178(4):365-371.

31. Ulberstad F, Boström H, Chavanon ML, et al. Objective measurement of attention deficit hyperactivity disorder symptoms outside the clinic using the QbCheck: Reliability and validity. *International journal of methods in psychiatric research.* 2020:e1822.

32. Bijlenga D, Ulberstad F, Thorell LB, Christiansen H, Hirsch O, Kooij JJS. Objective assessment of attention-deficit/hyperactivity disorder in older adults compared with controls using the QbTest. *International Journal of Geriatric Psychiatry.* 2019;34(10):1526-1533.

33. Groom MJ, Young Z, Hall CL, Gillott A, Hollis C. The incremental validity of a computerised assessment added to clinical rating scales to differentiate adult ADHD from autism spectrum disorder. *Psychiatry Research.* 2016;243:168-173.

34. Hirsch O, Christiansen H. Factorial Structure and Validity of the Quantified Behavior Test Plus (Qb+©). *Assessment.* 2017;24(8):1037-1049.

35. Reh V, Schmidt M, Lam L, et al. Behavioral Assessment of Core ADHD Symptoms Using the QbTest. *Journal of attention disorders.* 2015;19(12):1034-1045.

36. Caserta MT, Hall CB, Canfield RL, et al. Early developmental outcomes of children with congenital HHV-6 infection. *Pediatrics.* 2014;134(6):1111-1118.

37. Skotak M, LaValle C, Misistia A, Egnoto MJ, Chandra N, Kamimori G. Occupational blast wave exposure during multiday 0.50 caliber rifle course. *Frontiers in Neurology.* 2019;10(JUL).

38. Bijlenga D, Jasperse M, Gehlhaar SK, Sandra Kooij JJ. Objective QbTest and subjective evaluation of stimulant treatment in adult attention deficit-hyperactivity disorder. *European Psychiatry.* 2015;30(1):179-185.

39. Hall CL, Selby K, Guo B, Valentine AZ, Walker GM, Hollis C. Innovations in Practice: an objective measure of attention, impulsivity and activity reduces time to confirm attention deficit/hyperactivity disorder diagnosis in children – a completed audit cycle. *Child and Adolescent Mental Health.* 2016;21(3):175-178.

40. Tallberg P, Råstam M, Wenhov L, Eliasson G, Gustafsson P. Incremental clinical utility of continuous performance tests in childhood ADHD - an evidence-based assessment approach. *Scandinavian journal of psychology.* 2019;60(1):26-35.

41. Servatius RJ, Spiegler KM, Handy JD, Pang KCH, Tsao JW, Mazzola CA. Neurocognitive and Fine Motor Deficits in Asymptomatic Adolescents during the Subacute Period after Concussion. *Journal of Neurotrauma.* 2018;35(8):1008-1014.

42. Clark JF, Ellis JK, Burns TM, Childress JM, Divine JG. Analysis of Central and Peripheral Vision Reaction Times in Patients with Postconcussion Visual Dysfunction. *Clinical Journal of Sport Medicine.* 2017;27(5):457-461.

43. Hunzinger KJ, Sanders EW, Deal HE, et al. The use of a visual motor test to identify lingering deficits in concussed collegiate athletes. *J Clin Transl Res.* 2020;5(4):178-185.

44. Servatius RJ, Handy JD, Doria MJ, et al. Stress-Related Mental Health Symptoms in Coast Guard: Incidence, Vulnerability, and Neurocognitive Performance. *Front Psychol.* 2017;8:1513.

45. Spira JL, Lathan CE, Bleiberg J, Tsao JW. The impact of multiple concussions on emotional distress, post-concussive symptoms, and neurocognitive functioning in active duty United States marines independent of combat exposure or emotional distress. *J Neurotrauma.* 2014;31(22):1823-1834.

46. Blackwell C, Cary K, Holst K, et al. Dynavision Normative Data for Healthy Adults: Reaction Test Program. *The American journal of occupational therapy : official publication of the American Occupational Therapy Association.* 2020;74(1):7401185060p7401185061-7401185060p7401185066.

47. Mayes SD, Gordon M, Calhoun SL, Bixler EO. Long-Term Temporal Stability of Measured Inattention and Impulsivity in Typical and Referred Children. *J Atten Disord.* 2014;18:23-30.

48. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med.* 2016;15(2):155-163.

49. Qin S, Nelson L, McLeod L, Eremenco S, Coons SJ. Assessing test-retest reliability of patient-reported outcome measures using intraclass correlation coefficients: recommendations for selecting and documenting the analytical formula. *Qual Life Res.* 2019;28(4):1029-1033.

50. Fagan JF, 3rd, Singer LT, Montie JE, Shepherd PA. Selective screening device for the early detection of normal or delayed cognitive development in infants at risk for later mental retardation. *Pediatrics.* 1986;78(6):1021-1026.

51. Guzzetta A, Mazzotti S, Tinelli F, et al. Early assessment of visual information processing and neurological outcome in preterm infants. *Neuropediatrics.* 2006;37(5):278-285.

52. Greenberg LM, Kindschi CL, Dupuy TR, Hughes SJ. T.O.V.A. Clinical Manual. Test of variables of attention continuous performance test. 2007.

53. Forero CG. Cronbach's Alpha. In: Michalos AC, ed. *Encyclopedia of Quality of Life and Well-Being Research.* Dordrecht: Springer Netherlands; 2014:1357-1359.

Classification of Attention Task Performance Recorders
FDA Questions

Neurological Devices Panel of the Medical Devices Advisory Committee June 3-4, 2021

1. FDA has identified the following risks to health for attention task performance recorders intended to 1) measure reaction time and associated patient performance in response to attention tasks and 2) aid in assessment or diagnosis of specific diseases or conditions.

Risks to Health and Descriptions/Examples for Attention Task Performance Recorders Intended to Measure Reaction Time and Associated Patient Performance in Response to Attention Tasks, Without Aiding in Assessment or Diagnosis

Identified Risk	Description/Examples
Patient discomfort (e.g., visual or mental fatigue)	• Use of the devices can cause patient discomfort, such as visual or mental fatigue.
Incorrect or inaccurate measurements of reaction time or other attention tasks	• Use of the devices can result in incorrect or inaccurate measurements of reaction time or other attention tasks based on associated patient performance

Risks to Health and Descriptions/Examples for Attention Task Performance Recorders Intended to Aid in Assessment or Diagnosis of Specific Diseases or Conditions

Identified Risk	Description/Examples
Patient discomfort (e.g., visual or mental fatigue)	• Use of the devices can cause patient discomfort, such as visual or mental fatigue.
Incorrect or inaccurate results leading to inaccurate assessment or delayed diagnosis, both of which could result in inappropriate therapy or delay in treatment	• A false positive result means that the device indicates the patient has the clinical condition or disease of interest, such as ADHD or be at risk of cognitive impairment, when in fact none is present. • A false negative result means that the device indicates the patient does not have the clinical condition or disease of interest, such as ADHD or be at risk of cognitive impairment, when in fact the clinical condition or disease is present.

Please comment on whether you agree with inclusion of all the risks in the overall risk assessment of attention task performance recorders under product code "LQD". In addition, please comment on whether you believe that any additional risks should

be included in the overall risk assessment of these attention task performance recorders.

2. Section 513 of the Food, Drug, and Cosmetic Act states a device should be Class III if:

    - insufficient information exists to determine that general controls are sufficient to provide reasonable assurance of its safety and effectiveness or that application of special controls would provide such assurance, AND

    - if, in addition, the device is life-supporting or life-sustaining, or for a use which is of substantial importance in preventing impairment of human health, or if the device presents a potential unreasonable risk of illness or injury.

    A device should be Class II if:

    - general controls by themselves are insufficient to provide reasonable assurance of the safety and effectiveness, AND

    - there is sufficient information to establish special controls to provide such assurance.

    A device should be Class I if:

    - general controls are sufficient to provide reasonable assurance of the safety and effectiveness, OR
    - insufficient information exists to:
        - determine that general controls are sufficient to provide reasonable assurance of the safety and effectiveness, OR
        - establish special controls to provide such assurance, BUT
            I. is not purported or represented to be for a use in supporting or sustaining human life or for a use which is of substantial importance in preventing impairment of human health, and
            II. does not present a potential unreasonable risk of illness or injury.

    FDA believes general controls by themselves are insufficient to provide reasonable assurance of the safety and effectiveness and sufficient information exists to establish special controls to adequately mitigate the risks to health and provide reasonable assurance of device safety and effectiveness for this device type. As such, FDA believes that Class II is the appropriate classification for attention task performance recorders. Following are risk/mitigation tables, which outline the identified risks to health for this device type and the recommended controls to mitigate the identified risks, delineated by intended use:

Risk/mitigation recommendations for attention task performance recorders <u>intended to measure reaction time and associated patient performance in response to attention tasks only, without aiding in assessment or diagnosis</u>

Identified Risk	Recommended Mitigation Measure
Patient discomfort (e.g., visual or mental fatigue)	• Labeling
Incorrect or inaccurate measurements of reaction time or other attention tasks	• Non-clinical performance testing • Software verification, validation, and hazard analysis • Labeling

Risk/mitigation recommendations for attention task performance recorders <u>intended to measure reaction time and associated patient performance in response to attention tasks</u>

Identified Risk	Recommended Mitigation Measure
Patient discomfort (e.g., visual or mental fatigue)	• Labeling
Incorrect or inaccurate results leading to inaccurate assessment or delayed diagnosis, both of which could result in inappropriate therapy or delay in treatment	• Clinical performance testing • Non-clinical performance testing • Software verification, validation, and hazard analysis • Labeling

a. **Please discuss whether the identified special controls appropriately mitigate the identified risks to health for attention task performance recorders *intended to measure reaction time and associated patient performance in response to attention tasks only, without aiding in assessment or diagnosis*. Please also discuss whether additional or different special controls are recommended.**

1. The technical parameters of the device's hardware and software must be fully characterized and be accompanied by appropriate non-clinical testing:
    a. Hardware specifications must be provided. Appropriate verification, validation and hazard analysis must be performed, including applicable electrical safety testing.
    b. Software, including any proprietary algorithm(s) used by the device to measure reaction time and output other measures of attention, associated activities and related task performance, must be described in detail in the Software Requirements Specification (SRS) and Software Design Specification (SDS). Appropriate software verification, validation and hazard analysis must be performed.

2. Non-clinical device performance evaluation must demonstrate accurate and precise measurement of patient reaction times in response to task stimuli.

3. The labeling must include:
    a. A warning that the device is not intended to aid in patient assessment or diagnosis of specific diseases or conditions.
    b. Any instructions technicians must convey to patients regarding safe and effective administration of the specific tasks and collection of task performance data.

b. **Please discuss whether the identified special controls appropriately mitigate the identified risks to health for attention task performance recorders *intended to measure reaction time and associated patient performance in response to attention tasks for the aid in assessment or diagnosis of specific diseases or conditions*. Please also discuss whether additional or different special controls are recommended.**

1. Clinical device performance evaluation must validate that the device outputs accurately and precisely assess patient symptomology associated with the specific disease or condition for which the device is intended to assess or diagnose. The testing must:
    a. Evaluate agreement between device output and patient symptomology.
    b. Evaluate device test-retest reliability.
    c. Describe construction of any normative or reference database, which includes the following:
        i. How the clinical work-up was completed to define the reference population, including the establishment of inclusion and exclusion criteria.
        ii. Statistical methods and model assumptions used.

2. The technical parameters of the device's hardware and software must be fully characterized and be accompanied by appropriate non-clinical testing:
    a. Hardware specifications must be provided. Appropriate verification, validation and hazard analysis must be performed, including applicable electrical safety testing.
    b. Software, including any proprietary algorithm(s) used by the device to measure reaction time and output other measures of attention, associated activities and related task performance, must be described in detail in the Software Requirements Specification (SRS) and Software Design Specification (SDS). Appropriate software verification, validation and hazard analysis must be performed.

3. Non-clinical device performance evaluation must demonstrate accurate and precise measurement of patient reaction times in response to task stimuli.

4. The labeling must include:
    a. A summary of any clinical testing conducted to demonstrate that the device outputs accurately and precisely assess patient symptomology

associated with the specific disease or condition for which the device is intended to assess or diagnose. The summary of testing must include the following:

i. Agreement between device output and patient symptomology.
ii. Device test-retest reliability.
iii. A description of any normative or reference database, which includes the following:
    1. How the clinical work-up was completed to define the reference population, including the establishment of inclusion and exclusion criteria.
    2. How reference values will be reported to the user.
    3. Representative screenshots and reports that will be generated to provide the user results and reference data.
    4. Statistical methods and model assumptions used.
    5. Whether or not the database was adjusted due to differences in age, gender, or other factors.

b. A warning that the device is intended to aid in patient assessment or diagnosis by a trained physician and is not intended for stand-alone use.
c. Any instructions that technicians must convey to patients regarding safe and effective administration of the specific tasks and collection of task performance data.

3. **Please discuss whether you agree with FDA's proposed classification of Class II with special controls for attention task performance recorders. If you do not agree with FDA's proposed classification, please provide your rationale for recommending a different classification.**