Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms

Thomas D. Parsons\textsuperscript{a,b}, Timothy McMahan\textsuperscript{b} and Robert Kane\textsuperscript{c}

\textsuperscript{a}Department of Psychology, University of North Texas, Denton, TX, USA; \textsuperscript{b}Computational Neuropsychology and Simulation Laboratory, University of North Texas, Denton, TX, USA; \textsuperscript{c}Cognitive Consults and Technology LLC, Washington, DC, USA

ABSTRACT

Objective: Clinical neuropsychologists have long underutilized computer technologies for neuropsychological assessment. Given the rapid advances in technology (e.g. virtual reality; tablets; iPhones) and the increased accessibility in the past decade, there is an on-going need to identify optimal specifications for advanced technologies while minimizing potential sources of error. Herein, we discuss concerns raised by a joint American Academy of Clinical Neuropsychology/National Academy of Neuropsychology position paper. Moreover, we proffer parameters for the development and use of advanced technologies in neuropsychological assessments. Method: We aim to first describe software and hardware configurations that can impact a computerized neuropsychological assessment. This is followed by a description of best practices for developers and practicing neuropsychologists to minimize error in neuropsychological assessments using advanced technologies. We also discuss the relevance of weighing potential computer error in light of possible errors associated with traditional testing. Throughout there is an emphasis on the need for developers to provide bench test results for their software’s performance on various devices and minimum specifications (documented in manuals) for the hardware (e.g. computer, monitor, input devices) in the neuropsychologist’s practice. Conclusion: Advances in computerized assessment platforms offer both opportunities and challenges. The challenges can appear daunting but are a manageable and require informed consumers who can appreciate the issues and ask pertinent questions in evaluating their options.

1. Introduction

Although the past few decades have been notable for rapid technological advances in computer platforms (e.g. virtual reality, smartphones, tablets, Internet) and neuroinformatics (e.g. human genome, proteome, connectome projects, neuroimaging), clinical neuropsychological assessments are far from keeping pace (Rabin, Paolillo, & Barr, 2016). Thirty years
ago, Meehl (1987) called for clinical psychologists to embrace the technological advances then prevalent in our society. A decade after Meehl, Sternberg (1997) described the then current standardized tests (e.g. Wechsler scales) as differing little from tests that had been used throughout the twentieth century. Unfortunately, tests used in neuropsychology today remain closer to the now obsolete black and white televisions, vinyl records, and rotary-dial telephones than they are to the iPhone. Contemporary with Sternberg, Dodrill (1997) contended that neuropsychologists had made much less progress than would be expected in both absolute terms, and in comparison, with the progress made in other clinical neurosciences. According to Dodrill, the advances in neuropsychological assessment (e.g. Weschler scales) have resulted in new tests that are by no means conceptually or substantively better than the old ones.

Today’s technologies offer neuropsychologists the potential for both conceptual and substantive advances. Recently, some neuropsychologists have called for clinical neuropsychologists to embrace technological advances and argue that these technologies may transform neuropsychogy’s concepts and methods (Bigler, 2016; Bilder, 2011; Jagaroo, 2009; Parsons, 2016). Specifically, there is need for an incorporation of findings from the human genome project, advances in psychometric theory, information science, and advanced technologies (e.g. virtual reality-based assessments). A shift toward evidence-based science and praxes is possible if neuropsychologists understand the need for innovations in neuroinformatics (e.g. ontologies and collaborative knowledge bases) and the design of computerized assessment methods. Recent discussions have built on this emphasis and have emphasized the need for neuropsychologists to consider both the adoption and development of advanced technologies (e.g. American Academy of Clinical Neuropsychology’s Disruptive Technology Initiative; see also Kane & Parsons, 2017). One example is the development of virtual reality platforms as special cases of CNADS that offer increased ecological validity (Parsons, Carlew, Magtoto, & Stonecipher, 2017). In addition to these enhanced technologies, there is need for enhancing data logging and output. Many paper-and-pencil measures have limited (if any) data about an individual patient’s practice effects on tests like the Delis–Kaplan Executive Function System’s Color Word Interference task (and Switching for that matter). Such measures present multiple stimuli at once and miss out on the patient’s response to each stimulus, as well as the learning that occurs. The logging of this data for each patient response to each stimulus may allow for new paradigms.

Unfortunately, practicing neuropsychologists continue to be slow in adopting advances in technologies (Rabin et al., 2014, 2016). Instead of embracing the many advanced technologies available, most neuropsychologists go little further than an occasional use of computer-automations of paper-and-pencil assessments. Steps have been taken to discuss the potential advantages and disadvantages of these basic computer-automated assessments (Bauer et al., 2012; Cernich, Brennana, Barker, & Bleiberg, 2007; Parsey & Schmitter-Edgecombe, 2013; Schatz & Browndyke, 2002; see Table 1 for examples of computer-automated neuropsychological assessment batteries).

Optimism is apparent in reviews that laud computer automated neuropsychological assessments for their potential for augmenting test administration (Bauer et al., 2012), collection of normative data (Bilder, 2011), scoring (Woo, 2008), and interpreting results (Russell, 2000). Furthermore, computers have been noted for their enhanced (over paper-and-pencil administrations) presentation of complex stimuli (Gur et al., 2001; Schatz & Browndyke, 2002) and response logging (Crook, Kay, Larrabee, Grant, & Adams, 2009).
Table 1. Examples of computerized neuropsychological assessment batteries.

<table>
<thead>
<tr>
<th>Test</th>
<th>Approximate year released&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Platforms</th>
<th>Type</th>
<th>System requirements</th>
<th>Internet required</th>
<th>Input method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated Neuropsychological Assessment Metrics (ANAM)</td>
<td>1992</td>
<td>Windows 7 or newer</td>
<td>Text/2D Tests</td>
<td>2.5 GHz Intel Core i5-4200 M</td>
<td>No</td>
<td>Keyboard, Mouse</td>
</tr>
<tr>
<td>California Computerized Assessment Package (CALCAP)</td>
<td>1986</td>
<td>Windows 95/98</td>
<td>Text</td>
<td>Intel 80286 or faster</td>
<td>No</td>
<td>Keyboard</td>
</tr>
<tr>
<td>Cambridge Neuropsychological Test Automated Battery (CANTAB)</td>
<td>1985/2012</td>
<td>Apple iOS, Windows 7, Windows 8.1</td>
<td>Text/2D Tests</td>
<td>Apple 9.7&quot; iPad Pro, Windows touch screen devices</td>
<td>Not Required but is recommended</td>
<td>Finger touch</td>
</tr>
<tr>
<td>CNS Vital Signs</td>
<td>2004</td>
<td>Mac OS 10.6.8 or higher Internet Browsers</td>
<td>Text/2D Tests</td>
<td>2 GHz Pentium or better</td>
<td>Yes for the online version</td>
<td>Keyboard, Mouse</td>
</tr>
<tr>
<td>CogScreen</td>
<td>1995</td>
<td>Windows XP or newer; PC</td>
<td>Text/2D Tests</td>
<td>Any Machine capable of running Windows</td>
<td>No</td>
<td>Keyboard, stylus or touch Screen</td>
</tr>
<tr>
<td>ImPACT</td>
<td>1998/2007</td>
<td>PC/Mac with an internet browser</td>
<td>Text/2D Tests</td>
<td>Any Machine capable of running an internet browser</td>
<td>Yes</td>
<td>Keyboard, mouse</td>
</tr>
<tr>
<td>MicroCog</td>
<td>1993</td>
<td>Microsoft Windows 2000 or newer</td>
<td>Text/2D Tests</td>
<td>133 MHz Processor</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>MYBRAINTTEST</td>
<td>2010</td>
<td>PC/Mac with an internet browser</td>
<td>Text/2D Tests</td>
<td>Any Machine capable of running an internet browser</td>
<td>Yes</td>
<td>Keyboard, mouse</td>
</tr>
<tr>
<td>NeuroTrax Mindstreams</td>
<td>2000</td>
<td>Windows XP or newer</td>
<td>Text/2D Tests</td>
<td>1.4 GHz Processor</td>
<td>Yes</td>
<td>Keyboard, mouse</td>
</tr>
<tr>
<td>NIH Toolbox</td>
<td>2006</td>
<td>Windows 7 PC with an internet browser</td>
<td>Text/2D Tests/Real Life</td>
<td>Windows 7, Adobe Flash, and Java</td>
<td>Yes</td>
<td>Keyboard, mouse, various measuring devices</td>
</tr>
</tbody>
</table>

<sup>a</sup>The release year was an approximate date gleaned from first publication date of manuals and/or articles. For the NIH Toolbox, 2006 was the year that the original grant was funded.
Concerns have also been raised. For example, research needs to be done to assess whether the patient’s perception of and responses to the computer-generated stimuli are significantly different from their responses to traditional paper-and-pencil measures (Cernich et al., 2007). Furthermore, patients have differing levels of familiarity with computer interfaces (e.g. mouse, keyboard; Iverson, Brooks, Ashton, Johnson, & Gualtieri, 2009). The concern is that a computerized version of a task may produce a different result or assess a different capability than its paper-and-pencil counterpart. Each neuropsychologist must weigh the advantages and disadvantages of adding computerized assessments for their patients. Of specific concern are technical considerations and potential errors that can be found in computerized tests relative to hardware and software interactions (see Table 2).

The American Academy of Clinical Neuropsychology (AACN) and the National Academy of Neuropsychology (NAN) recently published a joint position paper on computerized neuropsychological assessment devices (CNADS). According to the consensus statement, CNADS were defined as ‘any instrument that utilizes a computer, digital tablet, hand-held device, or other digital interface instead of a human examiner to administer, score, or interpret tests of brain function and related factors relevant to questions of neurologic health and illness’ (Bauer et al., 2012, p. 2). It is important to note that their definition emphasized the administration of cognitive assessments by a computer ‘instead of’ a human examiner. Given today’s technologies, computers are not necessarily a replacement of the neuropsychologist. Instead, computerized neuropsychological platforms may be understood as tools used by a neuropsychologist that update paper-and-pencil paradigms and allow for enhanced stimulus presentation and logging of patient data (Parsons, 2016).

While, the clear majority of neuropsychological assessments still use paper-and-pencil-based technology, a recent survey of neuropsychologists revealed an increased likelihood of computerized test utilization among newer neuropsychologists (Rabin et al., 2014). This new generation of neuropsychologists, together with patients who expect twenty-first century health care to make full use of technology, underscore the need for renewed discussion of the promise and potential concerns related to computerized neuropsychological assessment and scoring technologies.

Given the rapid advances in technology and the increased accessibility, there is an ongoing need to identify optimal specifications, while minimizing potential sources of error. Herein, we discuss concerns raised by both Cernich et al. (2007) and the joint AACN and NAN position paper (Bauer et al., 2012). While, some of these issues are not as relevant given hardware advances, there are cases in which these concerns continue as advanced platforms (e.g. virtual reality, touch screen interfaces) become more prevalent. Moreover, we proffer parameters for the development and use of advanced technologies in neuropsychological assessments. First (Section 2), we attempt to put error associated with current computerized testing in the context of the error associated with traditional test administration. Next (Section 3), we describe software and hardware configurations that can impact a computerized neuropsychological assessment. While, some of these are less significant for traditional models of assessment, we emphasize the importance of these issues for advanced data analytics (e.g. neuroinformatics) and technologies (e.g. virtual reality) that move beyond current paradigms. Third (Section 4), we describe best practices for developers and practicing neuropsychologists to minimize error in both current technologies and emerging computerized neuropsychological assessment platforms. Throughout there is an emphasis on the need for developers to provide bench test results for their software’s performance on various
devices and minimum specifications (documented in manuals) for the hardware (e.g. computer, monitor, input devices) in the neuropsychologist’s practice.

2. Error in computerized vs. traditional paper-and-pencil measures

Computers have added both capabilities and challenges to the practice of neuropsychological assessment. The capabilities include, but are not limited to, the ability to capture more detailed

Table 2. Computer-based neuropsychological assessments: promise and concerns.

<table>
<thead>
<tr>
<th>Promise</th>
<th>Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Administration</strong> (stimulus presentation)</td>
<td><strong>Administration</strong></td>
</tr>
<tr>
<td>• Greater levels of control over administration and scoring</td>
<td>• Problematic hardware and software interactions can result in errors</td>
</tr>
<tr>
<td>• Less time needed for administration (2.5-2 h)</td>
<td>• A completely automated assessment does not allow for ‘testing of limits’;</td>
</tr>
<tr>
<td>• Improved precision of timing presentation</td>
<td>• A completely automated assessment does not allow for flexibility in evaluations</td>
</tr>
<tr>
<td>• Ability to implement a greater range of tasks (e.g. multi-tasking and divided attention)</td>
<td>• A completely automated assessment does not provide structured encouragement</td>
</tr>
<tr>
<td>• Randomization of test trials</td>
<td>• Can be administered without adequate neuropsychology training</td>
</tr>
<tr>
<td>• Adaptive testing protocols (reduced assessment times)</td>
<td></td>
</tr>
<tr>
<td>• Ease of developing alternate forms</td>
<td></td>
</tr>
<tr>
<td>• Accuracy of basal and ceiling levels</td>
<td></td>
</tr>
<tr>
<td>• Ability to discontinue testing relative to basal and ceiling levels</td>
<td></td>
</tr>
<tr>
<td>• Choice of multiple languages for administering tests in various languages</td>
<td></td>
</tr>
<tr>
<td>• Group administration</td>
<td></td>
</tr>
<tr>
<td>• Administration using portable devices (smartphones; handhelds)</td>
<td></td>
</tr>
<tr>
<td>• Administration procedure can be kept constant</td>
<td></td>
</tr>
<tr>
<td><strong>Scoring</strong> (Logging of Behavioral Responses)</td>
<td><strong>Scoring</strong> (Logging of Behavioral Responses)</td>
</tr>
<tr>
<td>• Enhanced logging capabilities allow for increased accuracy of measurement for response latency, efficiency, and response variance</td>
<td>• Patient responses to computerized test stimuli may not be identical to patient responses to paper-and-pencil presented stimuli</td>
</tr>
<tr>
<td>• Integrates and automate interpretive algorithms (e.g. decision rules for determining impairment)</td>
<td>• Computerized assessments may mask deficits that would otherwise be apparent in some populations (e.g. persons with autism may perform better on computerized tasks)</td>
</tr>
<tr>
<td>• Measure performance on time-sensitive tasks (e.g. reaction time)</td>
<td>• Computerized assessments may measure cognitive performance at a level that is seldom required in real-life situations</td>
</tr>
<tr>
<td><strong>Participant Experience</strong></td>
<td><strong>Participant Experience</strong></td>
</tr>
<tr>
<td>• Computerized assessments may be less daunting that responding a human examiner (may increase openness of respondents)</td>
<td>• Validity issues: Computerized measures may not be experientially (or psychometrically) equivalent to paper-and-pencil measures</td>
</tr>
<tr>
<td>• Computerized assessments may decrease the examiner’s influence upon the examination</td>
<td>• Some patients may have negative attitudes (including anxiety) about computers</td>
</tr>
<tr>
<td>• Computerized assessment may allow for greater accessibility/availability of neuropsychological services – especially in remote locations (teleneuropsychology)</td>
<td></td>
</tr>
<tr>
<td><strong>Normative Database</strong> (Neuroinformatics)</td>
<td><strong>Normative Database</strong> (Neuroinformatics)</td>
</tr>
<tr>
<td>• Computer automated logging allows for enhanced normative data collection and comparison</td>
<td>• Without a standardized approach to data logging, paper-and-pencil approaches will lag behind other disciplines like neuropsychiatry and behavioral neurology</td>
</tr>
<tr>
<td>• Computerized logging allows for straightforward and time efficient exporting of responses for data analytic purposes</td>
<td>• The challenges associated with using computers for assessment are amplified by the speed at which technology changes</td>
</tr>
<tr>
<td>• Computer-automated data exporting</td>
<td></td>
</tr>
</tbody>
</table>
response characteristics, the ability to implement a greater range of tasks including those involving multi-tasking and divided attention, and the ability to capture response times for multiple measures and to better integrate processing speed and efficiency into the assessment process. The challenges relate to complexities in understanding how computers work and how characteristics of computer systems affect test results. The challenges associated with using computers for assessment are amplified by the speed at which technology changes.

The error associated with computerized testing also must be evaluated in comparison to the error that exists when the neuropsychological evaluation is carried out by traditional means. Such errors may include errors in recording and scoring responses, in presenting stimuli at precise intervals, and in timing responses. The advent of computers has importantly prompted discussion and awareness of potential sources of error when using computers for assessment. Interestingly, there has been little research on sources of error associated with traditional paper-and-pencil testing. For example, there is a dearth of research on the variance in stopwatch presses both within a single test session (variance in a human pressing at the start and stop of a task) and between the neuropsychologists that use them (variance due to factors such as age, motor abilities). A stopwatch may be used to time a cognitive task that lasts exactly 10.00 s. Even a perfectly calibrated stopwatch, with 0.01 of a second resolution, is limited by the fact that it is operated manually by a human. If the human examiner’s reaction time varies by up to 0.1 s, then the stopwatch may be started 0.1 s after the event started and then stopped 0.1 s after it ends. This would mean that the neuropsychologist would record 10.20 when it actually took the patient 10 s to complete the task.

The same example can be applied to differences between the motor speed of various neuropsychologists. Sheppard and Vernon (2008) used data from 172 studies, with a total of 53,542 participants, to calculate effect sizes between groups (e.g. comparing young and older age participants) on speeded measures. They found that younger adults had shorter (i.e. faster) reaction time latencies than older adults. This means that there will likely be variance between stopwatch presses of a neuropsychologist at age 24 versus that neuropsychologist’s stopwatch pressing speed at age 70. If a neuropsychologist must flip stimulus cards every two seconds, there is a likelihood of variance in presentation rate both within a testing session and between examiners. The amount of variance and the significance of that variance has yet to be studied.

While, it is important to appreciate and to address limitations and error associated with computerized testing, clinicians must also evaluate this error in relationship to the goals of the assessment and the strengths and limitations of current approaches. If one thinks of computers and tablets as ways of streamlining and modernizing test administration, then the precision of electronic devices easily matches, and likely exceeds, that of a human examiner. Concerns about sources of error in computerized testing are largely driven by interface issues related to ways in which the patient interacts with the electronic device and by the desire to raise the testing bar in implementing new tasks, new paradigms such as virtual reality, and in capturing precise timing when measuring cognitive processing speed and efficiency.

3. Sources of error in computerized neuropsychological assessment

3.1. Hardware and software issues

Neuropsychological assessments performed using computerized platforms may be affected by the computer configuration (Bauer et al., 2012; Cernich et al., 2007). Various types of
software, processing power, and hardware may affect (1) presentation of stimuli; (2) behavioral responses, (3) data logging of performance; and (4) processing of behavioral data. While, there have been attempts to standardize aspects of computerized testing, variations still exist that could produce noise in computerized neuropsychological tests. The potential result of technological variance is unreliable data that deviates from normative data. For example, stimulus presentation and reaction times can be impacted by some hardware and software setups. While, most computerized neuropsychological assessments have taken efforts to control the precision and accuracy with which stimuli are presented as well as the response times, the technology used to present the stimuli can have a noticeable impact on performance and may be easily overlooked by neuropsychologists (MacInnes & Taylor, 2001; Mckinney, MacCormac, & Welsh-Bohmer, 1999). This is especially true for advanced platforms such as touch screens and 3D programs using virtual reality.

While, there are a number of validation studies underway to assess the psychometric properties of computerized neuropsychological assessments most of these software platforms run on standard personal computers that are not optimized for the accurate presentation of stimuli. Temporal uncertainty can be found in stimulus presentation durations caused by the underlying technologies and the default timing mechanisms found in computer operating systems. Plant and Turner (2009) provide the following formula related to how various factors can impact timings in computerized assessments for neuroscience research:

\[
\text{True time} = \text{Measured time} - (\text{Error due to paradigm features} + \text{Equipment error})
\]

For the neuropsychologist interested in computerized neuropsychological assessment, the expression (Error due to paradigm features + Equipment error) includes: (1) hardware timing latencies (from presentation devices, sound cards, and response devices); (2) driver issues and interactions (generic drivers that do not function in accord with the hardware manufacturer specifications and/or do not report correct timings); (3) coding errors that may occur inside recording software from hardware developers (unoptimized code, poorly scripted code, and incorrect settings); (4) operating system characteristics (variance in multi-tasking operating systems); (5) stimulus type (high-resolution audio/video presentation, rapid serial visual presentation); (6) software interactions (anti-virus checkers, anti-spyware packages, automatic updates and over-the-air push); (7) hardware interactions (e.g. issues in synchronizing hardware in neuroimaging, eye tracking); (8) configuration settings (volumes, luminance, and voice key thresholds); (9) unaccounted for factors; and (10) the specific computerized neuropsychological assessment used for neuropsychological testing (bugs in computerized neuropsychological assessment software program).

Neuropsychologists endeavor to develop and utilize measures that optimize precision and accuracy. In addition to timing issues, concerns are often raised about whether computerized neuropsychological assessment devices can achieve consistently valid and reliable assessments with off-the-shelf personal computers that are often developed to be affordable instead of optimized to the highest specifications. Can neuropsychologists trust the consistency, precision, and accuracy of computerized neuropsychological assessment devices? From our perspective, the answer is yes, with some qualifications. Developers of computerized neuropsychological assessment devices need to build and thoroughly test the presentation and response timings of their assessment platforms (Plant, 2016). Furthermore, it is imperative that they document results and minimum specifications in their manuals (Bauer et al., 2012; Cernich et al., 2007). The range of potential error related to different systems

\[
\text{True time} = \text{Measured time} - (\text{Error due to paradigm features} + \text{Equipment error})
\]
should be specified. Error that occurs within a limited range may or may not be consequential depending on the use of the test. It may also be possible to compensate for error that is consistent and unidirectional (error always lesser or greater by a specific amount).

3.2. Timing error in operating systems

Some computerized neuropsychological assessments are run on a general-purpose operating system (e.g. Microsoft Windows) that may impact the predictability of a given neuropsychological assessment task scheduling process. Moreover, the operating system can introduce timing error because other processes and operating system functions take processing time away from the neuropsychological testing application. This is especially true for older computer platforms using single-core processors (computers manufactured around and before 2008) that only allow for access to one program at a given time. This is less of an issue for newer computers because processor manufactures have been building multi-core processors since the late 2000s (Intel i5 or i7), that make it possible for multiple programs to access the processor at the same time. This reduces timing errors caused by the operating system. Furthermore, computerized neuropsychological assessment software developers can develop their software to eliminate delays by implementing priority boosting and frame pre-computation. It is important to note, however, that doubling up using a low latency operating system and specialized software (with built-in performance boosts) may result in poorer overall performance (suboptimal resource allocation).

It may be helpful for developers of computerized neuropsychological assessments to implement dedicated platforms using real-time operating systems (e.g. Linux 2.6.33-29-real time). Real-time operating systems differ from general-purpose operating systems found on most personal computers. General-purpose operating systems (e.g. Microsoft Windows) are unpredictable because they maximize throughput even if background tasks are neglected for long durations. Contrariwise, the scheduling in real-time operating systems emphasizes predictability over maximization of the overall throughput. Hence, real-time and low-latency operating systems may be preferable for clinical neuropsychological assessments that have high accuracy requirements. At minimum, future studies should be performed that examine whether computerized neuropsychological assessments running on a general-purpose operating system are as accurate as the same computerized neuropsychological assessment running on a real-time operating system (see Garaizar, Vadillo, López-de-Ipiña, & Matute, 2014 for an example of measuring software timing errors in cognitive neuroscience experiments). Again, the magnitude and predictability of the error are important considerations.

3.3. Hardware-based stimulus presentation issues

First, stimulus presentation can be impacted by an older monitor’s refresh rate (Cernich et al., 2007). While, cathode ray tubes (CRT) were a mainstay of display technology until around 2007, CRTs have largely been superseded by newer display technologies such as liquid-crystal display (LCD), light-emitting diode (LED), plasma display, and organic light-emitting diode (OLED) screens. Many of today’s monitors have a refresh rate of 60 hertz, which means that visual stimuli are displayed only at a rate of once every 16.677 milliseconds (ms), irrespective of when the computerized neuropsychological software designates that a
stimulus should be on the monitor screen. Given that the visual timing is discrete (not continuous) visual stimuli can only be presented at discrete timesteps. The issue for visual timing is synchronization with the monitor refresh rate. While, the majority of low cost (under $200.00) LCD and LED monitors currently in use and on the market, have a standard refresh rate of 60 Hz and a frame fresh rate of 10–18 ms, more expensive (over $200.00) LED monitors (e.g. those designed for playing videogames and non-immersive virtual environments) can have higher refresh rates up to 180 Hz (overclocked) which reduces the frame refresh down to 5–8 ms. It is important to note that in 2016, manufactures released monitors that have refresh rates of 240 Hz which reduces the frame refresh down to 4 ms.

In 2013, video card manufactures developed a new adaptive sync technology (NVidia G-SYNC, AMD – FREESYNC). This advanced synchronization technology allows a dedicated graphics card (inside the computer) to instruct the monitor when to refresh with subsequent frame data. Figure 1 illustrates how standard computers (without sync technology) draw frames before the monitor is ready to receive them.

Without sync technology, the monitor either tears the pre-mature draw frame (by stopping the previous draw and starting the drawing of new frames) which produces visual artifacts, or the monitor can wait and draw the frames in the next cycle which causes measurement error due to lag. Figure 1 illustrates the way in which sync technology fluctuates the refresh rate of the monitor to match the speed of the computer production of draw frames. The communication between the monitor and graphics card removes screen tearing and measurement error from lag.

A final note on stimulus presentation involves presentation of auditory stimuli. While a great deal of computerized neuropsychological assessment involves visually presented stimuli, clinical neuropsychologists are often faced with the need to use auditory stimuli

![Figure 1. Comparison of frame drawing in computers with and without synchronization.](image)

**Notes:** Computers without sync technology send generated draw frames as soon they are finished—regardless of whether the monitor is ready to receive them. If the monitor is not ready, then the new frame is drawn before the previous one finishes (tear arrow). One solution is to hold the draw frame until the monitor is ready. However, so doing introduces lag (lag arrow). Sync technology allows the computer to communicate with the monitor in a manner that synchronizes the refresh rate to the output of the graphics card. This alleviates screen tearing and lag.
instead of, or in addition to, visual stimuli. The sound cards used in some computers are hampered by start-up latency, in which a delay is present between the moment auditory stimuli are requested to be played and the time that the audio can be physically detected at the level of the speakers. The difference can be notable on laptops due to the unpredictable delay between clicking on the input hardware and receiving auditory feedback. It is important to note that sound cards often have some degree of start-up latency. Given that latencies can vary significantly among various sound cards and manufacturers rarely specify such timings, there is need in the neuropsychology community to research these timings. Although it is not enough to ascertain the delay through software alone, external chronometry can be used for more exact quantification (Plant & Quinlan, 2013). In situations where exact quantification is desired, latency is easily observable if one compares the original sound card used by the developer with the neuropsychologists own sound card. Manualized administration instructions should inform the clinical neuropsychologist about the specific audio hardware (e.g. headphones or loudspeakers) needed for optimal and validated presentation. As mentioned above, if timing of stimuli (in this case audio stimuli) is critical, then the manual should provide minimum requirements for the type of soundcard and speakers/headphones. Additional information should be presented in manuals related to volume calibration of the testing platform. While, dealing with variations in sound and video cards can seem daunting, from a testing standpoint, the key issue is often that of synchronization so that the timing of a response only begins once the stimulus has been presented.

### 3.4. Hardware response inputs

An additional area of concern is the hardware used to log patient responses. A computerized neuropsychological assessment may require the patient to use a mouse, keyboard, or touch screen to respond to neuropsychological stimuli. Most neuropsychologists are aware that mice, keyboards, and touch pads are available in a large variety of models and types. While, optimal behavioral response logging may be found using input hardware to which the patient is most accustomed, it is important to consider the potential limitations of various input devices. A significant amount of variability has been found in input registration among various models and types of mice (see Figure 2).

Plant, Hammond, and Whitehouse (2003) found statistically significant effects in a test of simple visual reaction times with a sample set of mice technologies tested alongside a standard keyboard. They found large variation between the mice tested in terms of minimum and maximum button-down latencies. For example, comparisons between the best and the worst performing mice revealed an absolute maximum difference of 66.9 ms and a minimum difference of 47.6 ms ($p < .001$). A significant was also found between the worst mouse and the standard keyboard ($p < .001$), with the standard keyboard proving significantly better than this mouse. A typical keyboard and mouse will have polling rates from 8 to 10 ms. This means that the computer only checks the mouse and/or keyboard for a patient’s response every 8–10 ms, regardless of when the patient performed the behavioral response.

Another issue that can occur when the computerized neuropsychological assessment requires the patient to press a button in rapid succession is that the contacts within the switch (key pressed) do not make contact cleanly. This inconsistent contact is called ‘bouncing.’ When the participant presses a key, a complete connection may not ensue. Although bouncing occurs in a matter of milliseconds, the microcontroller is moving so rapidly that a
transition will be detected each time the button bounces. This could be a major concern for a neuropsychological task that requires successive and rapid responses. In a study of bouncing time, Ganssle (2008) found bouncing times ranging from 10 to 157 ms among 18 hardware switch types. Neath and colleagues (2011) used a photodiode and custom hardware to examine presentation onset of visual stimuli and subsequent keyboard reaction times. Results revealed that keyboard reaction times could be as much as 100 ms too long, and that keyboards could vary by as much as 20 ms. It is important to note that the change in operating systems (from Windows XP to Windows 7) and changes in hardware (from PS/2 to USB) have allowed for faster polling rates (from 40 up to 1000 Hz) which potentially reduce the timing errors (from 25 to 1 ms).

An additional concern that has been voiced by neuropsychologists is that some touch screen monitors introduce timing error. This is apparent in the time it takes for the screen to detect and process where the participant touched the screen (Cernich et al., 2007). Until 2007, touch screens had very limited capabilities. For example, they had limited responsiveness to the patient’s touch (especially after a good deal of use) and at times only recognizing one touch at a time. The introduction of the Apple iPhone in 2007 marked an important advance in touch screen interfaces. Today, touch screens are greatly advanced with multi-touch capability and enhanced precision for finger location recognition. These advancements in touch screen technology have also reduced the latency that comes from translating

---

**Figure 2.** Comparison of mouse input delay error on unoptimized vs optimized hardware and software.

Notes: Input methods used for neuropsychological assessments must be optimized. When input devices are not properly synchronized with other hardware variance in timing may be introduced. An example of this is mouse input and monitor timing. If a low polling rate is used with a low monitor refresh rate there is potential for a change on the monitor screen to not be recognized for 16 ms. Utilizing a higher polling rate allows for input detection to occur more frequently. This allows for greater matching to a monitor that is refreshing more often. Having unoptimized software can lead to longer processing times in the central processing unit (CPU) and the graphics processing unit (GPU).
the patient’s touch to the corresponding function in the application. In 2012, the average response time was 100 ms for touch screens (e.g. when a patient drew a line on the touch screen the finger touch would be 100 ms ahead of the line being drawn). In that same year, Microsoft presented a touch screen which had a one millisecond response time that is equivalent to near real-time (http://phys.org/news/2012-03-microsoft-finger-1ms-touchscreen-video.html). In 2016, the touch screen latency for mobile devices (e.g. smartphones and tablets) was reduced to the 20 ms range (https://www.reddit.com/r/Android/comments/4emced/touch_latency_is_being_highly_disregarded/).

While, voice keys are not widely used in computerized neuropsychological assessment devices (most focus on button presses) future advancements in computerized assessment will also record vocal responses that may be assessed for accuracy and reaction time. Voice keys tend to function via the crossing of thresholds that can be understood as the number of times that an analog signal reaches a certain threshold (i.e. volume). It will be important that developers of computerized neuropsychological assessments convey information related to whether plosives, fricatives, and noise sounds (voiced or unvoiced) are present (Kessler, Treiman, & Mullennix, 2002; Tyler, Tyler, & Burnham, 2005). Furthermore, neuropsychologists will want to know what the thresholds are and whether settings are the same from one assessment to another. As with other response devices discussed in this manuscript, the hardware and interfaces utilized can have considerable variability. Computerized neuropsychological assessment developers need to ensure the calibration of voice keys and include tools that the neuropsychologist can use to validate this calibration.

When using wireless peripherals, it is important to minimize the distance between the device and its receiver. For example, moving a device from 10 m to within 1 m will decrease the potential of radio waves to hinder signal strength. Wireless networks, smart phones (and other wireless devices) can dilute signal strength of wireless peripherals. A simple fix is to remove unnecessary wireless devices when using wireless peripherals during neuropsychological testing. In addition to timing issues, various mice can have buttons in different locations and keyboards may vary in key position as well as the position of the board itself (e.g. laptop keyboards are often different from keyboards used for personal computers). Hence, it is important for clinical neuropsychologists to find out from the developers of computerized neuropsychological assessments what input hardware was used during validation. If clinical neuropsychologists choose to deviate from the input hardware used in standardization, steps should be taken to study the timing characteristics of the new input hardware so that recorded times can be adjusted. Without psychometric assessment of the reliability and validity of the new input hardware, input system functionality may introduce adverse effects.

3.5 Automatic updates and over-the-air push

An additional issue that may impact performance of computerized neuropsychological assessments is the automatic updates (bug fixes and patches) that personal computers, Macs, and Linux boxes receive to their operating systems. The impact of an update can easily go unnoticed. Computers found in different areas of a hospital or neuropsychologist’s practice may be on different revisions of the same operating system. As a result, they may have different timing characteristics. It is suggested that users check with the developers before installing updates to verify that the testing software have been properly tested with the
latest updates to the operating system. If it is determined that updates may cause adverse effects, neuropsychologists could consider running the computers that are used for neuropsychological assessments off network and not applying updates. These steps will help ensure consistency. It is important to note, however, that choosing to take a computer off network may increase security risk once the computer is reconnected to the network. This is because the machine may not receive important updates and/or it may not be running the latest security patches. Neuropsychologists should consult their IT department, the test developers, and the manual about these risks before deciding to go off network and/or no longer accept updates.

### 3.6. Tablets and phones

Developers of computerized neuropsychological assessments are beginning to move to unproven (for clinical neuropsychological assessment) platforms like Android tablets and Apple iPads. Tablets and smartphones are developed with a variety of hardware specifications (different processors, memory, display technology, and screen resolutions). Furthermore, they are used increasingly for virtual reality platforms (see Table 3). Like PC operating systems, Android and Apple operating systems allow background processes and applications to run that can use up valuable resources. This increases the potential for collecting data that contains errors. There are now a number of studies involving applications for iPad and Android tablet-based assessments of neurocognitive constructs. Unfortunately, only a handful of studies have validated the data acquired from a tablet device against traditional measures (Clionsky & Clionsky, 2014; Onoda et al., 2013; Rao et al., 2017). Others validated the tablet version of the neuropsychological assessment against a computer-based version (Black et al., 2013; Burke et al., 2016; Dorr, Lesmes, Lu, & Bex, 2013), or they simply limited their methods to only a tablet-based version with no criterion measure (Bertucco & Sanger, 2014; Van Tasell & Folkeard, 2013; Zhang, Red, Lin, Patel, & Sereno, 2013). In addition to these individual assessments, a number of traditional neuropsychological assessment batteries (Wechsler Adult Intelligence Scale – 4th Edition, Wechsler Intelligence Scale for Children – 4th Edition, Delis-Kaplan Executive Function System, and California Verbal Learning Test – 2nd Edition and Children’s Editions) are now available for iPad-based assessment, using ‘Q-interactive’ – an iPad application that allows neuropsychologists to administer clinical assessments using two tablets that are connected via Bluetooth technology (Pearson, 2013). It is important to note that some computerized assessments (e.g. Q-interactive) may not be interested in highly accurate response times. Instead, they may be interested in streamlining the administration of traditional paper-and-pencil tests. In such cases, issues of millisecond timing will be less important.

Unfortunately, there is very little discussion of the extent to which tablet and iPad-based neuropsychological assessments offer reliable assessments. Plant and Quinlan (2013) point to research suggesting that presentation, synchronization, and response timing errors in the hundreds of milliseconds are apparent. Apple has at least five iPad models with various versions of iOS deployed on each. Moreover, Apple has removed real-time mode on some iOS versions and shifted other features to the Core Audio functions. While this may enhance multitasking, and save battery life for everyday use, it can also impact neuropsychological assessment precision and accuracy. Plant and Quinlan point to a Microsoft Applied Sciences Group video that illustrates the typical 100-ms lag inherent in generic touch screens.
Table 3. Current popular virtual reality system specifications and hardware requirements.

<table>
<thead>
<tr>
<th>System</th>
<th>Hardware</th>
<th>Graphics card</th>
<th>Operating system</th>
<th>Sensors</th>
<th>Tracking</th>
<th>User interaction</th>
<th>Resolution</th>
<th>Refresh rate</th>
<th>Field of view</th>
<th>Connection</th>
<th>Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Daydream</td>
<td>Google Pixel, Huawei Mate 9 Pro, ZTE Axon 7, Motorola Moto Z</td>
<td>Phone Dependent</td>
<td>Android</td>
<td>Gyroscope, accelerometer, proximity sensor</td>
<td>Head tracking</td>
<td>Controller, Gamepad</td>
<td>Phone Dependent</td>
<td>90 Hz</td>
<td>96 degrees</td>
<td>Wireless</td>
<td>~10 ms</td>
</tr>
<tr>
<td>HTC VIVE</td>
<td>Intel i5-4590/AMD FX 8350, 4 GB of RAM</td>
<td>NVIDIA GeForce GTX 970/ AMD Radeon R9 290</td>
<td>Windows 7 or newer</td>
<td>Accelerometer, gyroscope, laser position sensor</td>
<td>Head tracking, room tracking</td>
<td>Controllers</td>
<td>2160x1200</td>
<td>90 Hz</td>
<td>110 degrees</td>
<td>Wired</td>
<td>~11 ms</td>
</tr>
<tr>
<td>Oculus Rift</td>
<td>Intel i3-6100/AMD FX 4350, 8 GB RAM</td>
<td>NVIDIA GeForce GTX 970/ AMD Radeon R9 290</td>
<td>Windows 7 or newer</td>
<td>Gyroscope, accelerometer, optical sensor</td>
<td>Head tracking, room tracking</td>
<td>Controller, Gamepad</td>
<td>2160x1200</td>
<td>90 Hz</td>
<td>110 degrees</td>
<td>Wired</td>
<td>~13 ms</td>
</tr>
<tr>
<td>Samsung Gear VR</td>
<td>Samsung phones: Galaxy S8, S8+, Galaxy S7, S7 Edge, S6 Edge+, S6, Galaxy S6 Edge, Note 5</td>
<td>Phone Dependent</td>
<td>Android</td>
<td>Gyroscope, accelerometer, proximity sensor</td>
<td>Head tracking</td>
<td>Gear Controller, Gamepad, Navigation pad</td>
<td>2560x1440</td>
<td>60 Hz</td>
<td>101 degrees</td>
<td>Wireless</td>
<td>~20 ms</td>
</tr>
</tbody>
</table>
They contend that an implication of the lag in generic touch screens is that reaction times in a neuropsychological assessment that rely on touch will have error relative to the latency of the touch registration on a given device.

In a recent assessment of timing delays between stimulus presentation and (simulated) touch response on iOS devices (3rd- and 4th-generation Apple iPads) and Android devices (Kindle Fire, Google Nexus, and Samsung Galaxy) at response intervals of 100, 250, 500, and 1000 ms, Schatz, Ybarra, and Leitner (2015) found significantly greater timing error on Android devices (Google Nexus and Samsung tablets; 81–97 ms) than found on Apple iPads (27–33 ms). Comparisons of the Apple devices, revealed that the iOS 7 had significantly lower timing error than the iOS 6. They also found that simple reaction time trials (250 ms) on tablet devices represent 12–40% error (30–100 ms) and decreased choice reaction time trials (3–5% error at 1000 ms). The inherent error suggests that neuropsychologists should take care in their reliance on tablet devices for accurately recording the initiation of the touch screen.

3.7. Internet-based neuropsychological assessment

Internet-based neuropsychological testing is another assessment modality that may be affected by participants’ computer configuration and Internet connection (Bauer et al., 2012; Feenstra, Vermeulen, Murre, & Schagen, 2016; Germine et al., 2012; Reips, 2002a). In addition to the hardware and software issues above, Internet-based neuropsychological assessment has specific challenges related to server-side and client side processing. Most Internet-based neuropsychological assessments involve a combination of server-side and client-side processing. For server-side processing, data are fetched from a database and a web page is created in the browser. For client-side (i.e. browser) processing, dynamic web pages respond directly to the patient’s behavioral responses without consulting the server. For Internet-based neuropsychological assessments that use server-side processing, the presentation of stimuli may be impacted by the quality and speed of the patient’s Internet connection. For these server-side applications, developers have designed browser plug-ins to run the script on the client side (see Figure 3).

While, older Internet-based neuropsychological assessments may have used server-side processing, advances in browsers, hardware, and more uniform standards for client-side computer languages (JavaScript, HTML, and CSS) have prompted a move toward Internet-based assessments using the client side processing. This shift limits the impact of technical variance. In fact, stimulus presentation with HTML5 has been shown to have the capacity for approaching millisecond accuracy (Garaizar et al., 2014). It is important to note that these approaches (server-side and client-side) can be combined to optimize an Internet-based neuropsychological measure’s reliability and usability, this is commonly referred to as store and forward approach. These newer approaches allow for optimal stimulus timing (client-side) with safe data transfer (server-side) to the database. Using this approach, Internet-based neuropsychological testing (using contemporary browsers) may approximate closely the performance of computerized neuropsychological assessment on local computers. It is important to note, however, that these performance standards drop when the Internet-based neuropsychological assessment makes extreme demands on the hardware (e.g. rapid, high-definition 3D graphics).
3.8. Data security

Data generated from computerized test measures is protected health information that should be stored on an encrypted hard drive or on an encrypted partition of the computer’s hard drive. The computer must also be password protected with access being granted to properly vetted personnel. Neuropsychologist using computerized tests are ultimately responsible for making sure that patient data is secure based on the information they are gathering. Test developers have the responsibility of documenting how they handle data and whether their test system properly produces encrypted data files with at least 128-bit encryption. Internet testing presents a specific challenge if patient identified test data are kept in the cloud for the clinician to download. Data should be encrypted from the moment it is collected and only be decrypted after it has been transmitted back to the neuropsychologist. However, the data is only available after the user (e.g. patient; participant) has finished the test. Store-and-forward fixes both of these issues by storing data locally until an uncritical time in the assessment when it then forwards the data to the server.

Figure 3. Comparison of server-side processing vs. client side processing. 

Notes: Assessments that utilize server-side processing are more susceptible to timing delays that result from the internet connection of the user (e.g. patient; participant) and server. The amount of time taken to send data to the server depends upon several factors including: the user’s (e.g. patient’s; participant’s) internet speed, the server’s internet speed, and the number of current connections to the server. Utilizing client-side processing allows for the removal of any error that may be introduced due to internet connections or busy servers. However, the data is only available after the user (e.g. patient; participant) has finished the test. Store-and-forward fixes both of these issues by storing data locally until an uncritical time in the assessment when it then forwards the data to the server.
advantage of large scale data aggregation presents complex security issues that must be addressed by the participating parties. Terms for such arrangements would have to be worked out in advance with only select information available for analysis and it may also require patient informed consent to participate in such a process.

4. Optimizing computerized neuropsychological assessments

While there are a number of concerns related to test software and hardware configurations for neuropsychological assessment, there are steps that can be taken to optimize computerized neuropsychological assessments. In this section, we discuss criteria that we view as important for making computerized neuropsychological assessments reliable and valid.

4.1. Testing hardware and software

Practicing neuropsychologists may use a variety of processors, stimulus presentations (e.g. monitors and audio systems), input devices, and types of machines (desktop computer, laptop, and tablet). Moreover, the computers in neuropsychologists’ offices will have varying amounts of memory installed. Developers of computerized neuropsychological assessments should test software on several systems, including older, lower-end hardware, such as those found in some practicing neuropsychologist’s offices. While a neuropsychologist may think that it is enough to purchase a computer with high processing power, it is also important that they have adequate memory and a graphics card that corresponds to the recommendation from the developers. If computers have too little memory, neuropsychologists may find that their machine is running slowly. In addition to adding RAM (random access memory) to their machines, neuropsychologists can close applications other than the computerized neuropsychological assessment before starting the test. Computer manufactures will commonly put unwanted applications (bloatware) on new computers that may use up valuable resources. When practicing neuropsychologists purchase computers from manufacturers it is important to make sure that any factory installed bloatware be removed. Along with adequate memory, a dedicated graphics card can alleviate strain from both the processor and memory by handling information that is being sent to the monitor screen. Selecting a proper graphics card is best done by determining future assessment requirements. For example, a text-based assessment would not have the same requirements as a virtual reality-based assessment. For current virtual reality platforms that use head mounted displays like the Oculus Rift and HTC VIVE, have minimum requirements for Nvidia GeForce GTX 970 graphics cards (See Table 3) that allow for the presentation of both enhanced 3D graphics and simpler 2D assessment text-based assessments.

For the neuropsychologists, it is important to note the developer’s minimum specifications. Developers should indicate in the manual what type of graphics card is recommended for the assessment. Again, the large variability in input registration that exists among different types and models of keyboards and mice means that the recommended peripherals need to be thoroughly tested by the developer. Neuropsychologists should consult manuals to ensure that they are using the optimal devices.

Temporal uncertainty is an important concern for developers of computerized neuropsychological assessment devices that attempt to assess precise reaction time. There are methods to determine the precision with which response times are captured and it is the
responsibility of the test developer to measure the accuracy of their response times, to be specific in their hardware requirements, and to provide clear and adequate documentation in their test manuals. Fortunately, comprehensive hardware–software systems are available for determining the bias and noise in any computerized neuropsychological assessment. For example, the Black Box Toolkit (www.blackboxtoolkit.com) and StimTracker (www.cedrus.com) provide the developer and the neuropsychologist external chronometry to assess millisecond timing accuracy. Both approaches are used in hundreds of laboratories around the world. It is important to note that assessing timing without specialist turnkey solutions requires the use of oscilloscopes, function generators, logic analyzers, relevant electronics, and engineering expertise. For Plant, Hammond, and Whitehouse (2002), the optimal (if only) way to satisfactorily address temporal uncertainty is through empirical and ethologically valid determination of the magnitude of the timing error via external chronometry. Response simulators like the Blackbox Toolkit simulate a response to a stimulus (e.g. a button press). The rate of responses can be programmed by the developer to generate a stereotypical pattern of reaction times. The response simulator can be used to assess whether the observed/recorded reaction times are consistent with the ‘true’ reaction times that are simulated by the response simulator (Plant, 2016; Plant & Quinlan, 2013).

Timing analysis tools like the Blackbox Toolkit include a piece of hardware that developers of computerized neuropsychological assessments should use to record stimulus presentation timings and respond with known response times. These timing analysis tools can detect and record visual and auditory stimulus onset and offset with sub-millisecond accuracy. They use opto-detectors attached to the computer monitor, and microphones that are positioned in front of a computer’s speakers. Signal cables can be connected to mouse and/or keyboard switches on the computer running the computerized neuropsychological assessment. Events with precise latencies (e.g. sounds and switch closures) can be generated. For example, the developer can use these devices to detect behavioral responses (e.g. button presses), as well as detect color changes on a monitor screen and/or audio feedback. For example, a keyboard or mouse can be dismantled and wires can be attached on either side of a key or mouse button. Then a switch closure can produce an event (e.g. keypress; mouse click) so that the timing analysis tool can generate responses to stimuli with a known, precise latency (e.g. keypress exactly 300 ms after the stimulus appears on a monitor screen). The timing analysis tool can assess the apparent response time, allowing the developer to identify the extent to which it diverges from the actual response time. As a result, the developer can identify the internal latencies of the computerized neuropsychological assessment before it ever reaches the practicing neuropsychologist’s office. As previously stated, it is imperative that the developer document in detail the results gleaned from assessment with various devices and recommend the optimal hardware setups for running the computerized neuropsychological assessments in the clinician’s office.

In addition to external chronometry, Salmon et al. (2016) discuss histogram plotting and modulo-tables with modulo-binning scores. While, developers need to use external chronometry before releasing their software (validation results should be provided in the manual), researchers and clinicians may want to do their own validation of existing computerized neuropsychological assessments using histograms and modulo-tables that only involve access to software analysis tools like CRAN-R that are freely available. Neuropsychologists can use statistical software (e.g. R, SAS, SPSS, Statistica) to develop histograms for visualizing the ranges in which reaction times are more or less common. On the x-axis they can plot
equal bins and on the y-axis they can plot the frequency of the reaction times. While histograms can be helpful for visualizing the accuracy of reaction times, this approach requires a fair amount of data. As a result, histograms from a small sample may not proffer an adequate amount of trials to visualize either systematic oscillations or smooth distributions of reaction times.

Salmon et al. (2016) discuss another approach for visualization of reaction time data that can be used with smaller sample sizes. This modulo-operation method involves taking all the reaction times and dividing them by 2, while also counting the number of times the remainder is 0 instead of 1. Next, the neuropsychologist would take all the reaction times and dividing them by 3, while counting how many times the remainder is 0, versus 1, versus 2. This process is repeated by dividing each reaction time by 4, then 5, then 6, and so on, while counting the remainders each time. According to Salmon et al. (2016), reaction times that are biased toward the refresh rates of the hardware being used will have reaction times that are divisible by values for which the refresh rate is divisible. This will result in non-uniform groupings of remainders. For example, a 60 Hz computer refreshes at a rate of 16 ms, so the neuropsychologist would divide all reaction times by values up to 16 ms. Next, the results can be placed in a modulo-table. Each row indicates the division number of choice (i.e. number by which each reaction time is being divided). Each column indicates count of the number of reaction times with a specific remainder (‘0’, ‘1’, ‘2’,). Although the numbers in the modulo-table are not enough for detection of oscillating patterns in the data, a conditional formatting function (available in Microsoft Excel) can be used for selecting color ranges to color code the cells in the table. If measurement of reaction time is free of noise, then the neuropsychologist will not see any reoccurring or clustering of color because the number of items for each divisor for each remainder should be approximately equivalent. When the neuropsychologist applies conditional formatting to color the table, the cells with the highest divisor would be the same color, with varying shades as the divisor. Furthermore, unlike the histogram approach, the modulo-table approach can also reveal non-uniform remainder groupings in small sample sizes. While, the tools discussed provide the practicing neuropsychologist ways to check timing resolution, the main burden is on the test developer to adequately vet their test systems and to provide detailed data pertaining to technical issues in test manuals.

4.2. Communication of software and hardware requirements

One of the first things that developers of computerized neuropsychological assessments can do to help the practicing neuropsychologist is to communicate the minimum requirements for the type and version of the operating system that are needed for the neuropsychology tests (see Table 4). Developers should also communicate minimum hardware specifications, as well as minimum processor and memory (RAM) requirements. Given that significant variability in input registration exists among different types and models of keyboards and mice (with statistically significant effects on reaction time logging; Plant et al. (2003), the input hardware used during development and testing should always be communicated. Developers can also aid practicing neuropsychologists through the incorporation of instructions that inform the neuropsychologist about specific audio hardware (e.g. monitors, headphones, or loudspeakers). Likewise, minimum requirements on type of soundcard should be provided.
Table 4. Recommended content of the user manual for computerized neuropsychological assessments.

System requirements

• Document the minimal and optimal computer system requirements:
  ◦ Processor; Memory; Hard Drive Space; Video Card; Operating System (Windows/Mac, 32/64 bit); Other Software requirements (Direct X, .Net version, database etc.)
  ◦ Documentation of compatible hardware and software combinations should include verified timing accuracy
• Document requirements for peripherals
  ◦ Recommended mouse and keyboard to use for minimization of measurement error
  ◦ Documents of compatible input devices should be included and their timing resolutions identified
• Include documentation regarding timing resolution, possible sources of error, and instructions for standard implementation
• Document the recommended screen resolution
  ◦ If required, included the recommended internet connection speed
• Document any special hardware that may be required

Installation

• Developers should do their best to avoid using third party applications as they increase the risk of measurement error
• If special hardware is required include instruction to install the device drivers
• Always document the prerequisite software that is required to run the neuropsychological assessments:
  ◦ List these in the order in which they must be installed
  ◦ Note which version of the prerequisite software is required
  ◦ Provide the prerequisite software or a link to the location of the software
• Provide all detail necessary for successful installation of the program:
  ◦ Include screen shots of the installation screens
  ◦ Include a troubleshooting list for any errors that the user may experience
  ◦ Provide any tips that may make installation easier for the user

Administration

• Document optimal test conditions
• Description of how to start neuropsychological assessments
• Provide details about the options that may be available in the neuropsychological assessments:
  ◦ Changing stimuli
  ◦ Connection to special hardware
  ◦ Changing display settings
• Include screen shots of the settings menus
• Include screen shots of the program running
• Provide tips or notes about the program:
  ◦ Keyboard commands
  ◦ Exiting the assessment
  ◦ Special instructions that the participant must follow to get accurate results

Accessing data

• Data registration
  ◦ Explain how the program stores data as the user goes through the assessment
  ◦ Provide a list of variables and description of variables that the assessment collects
  ◦ Describe safeguards against overwriting of existing data
  ◦ Test scores calculated
  ◦ Compound scores calculated
• Data storage
  ◦ Provide the location in which the data is stored
  ◦ Describe how to access the data
• If stored in database provide the following:
  ◦ How to access the database
  ◦ Login information for the database
  ◦ How to export the data from the database
  ◦ Security, level of encryption, and HIPAA compliance of database
  ◦ Information on obtaining a Business Associates Agreement for protected health data stored in separately from the user’s computer
• Additional issues:
  ◦ Let the examiner know if the participant must complete the assessment before it saves the data

(Continued)
We also suggest that developers document timing validation measures in their manuals, as a benchmark of good practice. From the point of view of clinical neuropsychologists, this level of documentation is imperative to ensure that their patients are receiving the best possible assessments. Furthermore, clinical neuropsychologists should be very careful about using any test in their practice that does not have such documentation.

4.3. Optimizing Internet-based neuropsychological assessment platforms

In addition to the above, Internet-based testing platforms have some unique safeguards that should be followed. First, developers should test various browser versions across operating systems to rule out bugs and to make sure that stimulus presentation is consistent (Reips, 2002b). Also, since browsers often receive regular (automatic) updates, developers should regularly test browser compatibility and make any needed program adaptations. Minimum requirements for type and version of browser software should be communicated to the practicing neuropsychologist. It is important to note that a well-designed platform will include browser and operating system detection that allows for the identification of the type of software the patient is using. Developers should regularly update neuropsychologists about problems with the software. Furthermore, developers should communicate actions that the neuropsychologist can take to remedy known issues.

Quality assurance for Internet-based neuropsychological assessments includes: (1) analysis of the program for functionality on all major operating systems, devices, and browsers; (2) communication in the manual technical minimum requirements (Bauer et al., 2012; Birnbaum, 2004; Cernich et al., 2007; Feenstra et al., 2016); and (3) HIPAA compliance for any patient data stored on cloud-based servers and rather than on the neuropsychologists own computer. Online cross-system and cross-browser testing services are available (e.g. Sauce Labs: https://saucelabs.com/ and Browserstack: https://www.browserstack.com/) for remote analytics of Internet-based neuropsychological assessments. These testing services can assess assorted types and versions of browsers. In addition to manual testing performed by
the developer, researcher, or clinician, the Internet-based neuropsychological assessment can be designed for automated cross-system and cross-browser testing.

While, there is some optimism about Internet-based neuropsychological assessments using Adobe Flash, HTML 5, and JavaScripting, caution may be warranted. According to Reimers and Stewart (2007, 2008) using Adobe Flash to collect precision timed behavioral responses with uncontrolled machines (used outside of the laboratory) resulted in 20 ms slower (on average) performance when compared to standardized machines used inside the lab. Interestingly, even these standardized machines tended to be approximately 10 ms slower than a calibrated Linux-based system.

4.4. Testing of touchscreen-based assessments

Developers of computerized neuropsychological assessments are beginning to move to platforms like Android tablets and Apple iPads. The touchscreen technology of these devices requires extensive timing analysis. Developers should validate timing accuracy of visual, audio and haptic elements using tools that allow for deep analysis of a touch-screen user interface. For example, Watchdog (from OptoFidelity, 2014) records the touchscreen with a video camera and uses software to automatically detect and report changes on the touchscreen with accurate timestamps. It also uses a pressure sensitive switch for touch detection. Again, developers should document in detail the results gleaned from assessment.

5. Conclusions

Computers can affect test administration at different levels from streamlining and modernizing some traditional approaches to presenting different types of tasks and capturing additional and important aspects of performance. Challenges in computerized testing come mostly from raising the bar and attempting to engage in more complex assessment such as implanting virtual reality and Internet-based testing and capturing different aspects of performance including response times to enhance the assessment of processing speed. In this paper, we discussed the ways in which the technology used to run computerized neuropsychological assessment software and the subsequent presentation of stimuli can have a noticeable impact on performance and may be easily overlooked by neuropsychologists. The issues addressed in this paper underscore the need to require publishers of computerized tests to address technical issues in their manuals (as outlined in Table 4). For traditional measures, technical manuals have been mostly statistical manuals providing detailed psychometric information. This type of psychometric data remains important for all assessment instruments, traditional and computerized. Computerized test manuals, however, must also provide additional technical information related to the level of response timing the test or test system claims to achieve and how this level of precision was verified. System requirements must be specific and as technology changes publishers must provide information on how these changes affect the performance of their software.

Although computerized neuropsychological assessment is open to errors from sub-optimal software and hardware configurations, developers of computerized neuropsychological assessments can limit these errors with external chronometry and the results of these studies should be included in test manuals. When developing tests for use with off-the-shelf computers, developers should make explicit attempts to exhaustively document compatible
hardware and software combinations with verified timing accuracy. While, it is not realistic to expect all neuropsychologists to carry out their own technical evaluations, clinical neuropsychologists must familiarize themselves thoroughly with test documentation to ensure that the hardware and software in their offices have the minimum specifications (documented in manuals). Increased utilization of computerized neuropsychological assessments may require enhanced training at various levels of the professional neuropsychologist’s development (Rabin et al., 2014). Furthermore, independent studies are encouraged that review technical data and that verify claims made by publishers. Some clinicians and researchers may also opt to use visualization methods (e.g. histograms, modulo-operation method; Salmon et al., 2016) to alleviate temporal uncertainty. Providing needed information to users about areas of potential concern, along with enhanced test manuals, and independent verification will permit neuropsychologists to make informed decisions in selecting which computerized neuropsychological assessment instruments best fit their clinical needs.

It is also important that the potential of computerized testing not be lost in concerns about sources of error. Paper-and-pencil tests are typically limited to measurement of total time on task. Computerized tasks/batteries provide the neuropsychologist with discrete measurement of reaction time for single attempts at a task. This offers promise for detection of specific deficits in inhibition, inattention, vigilance, or initiation of a task. Computerized assessments provide enhanced access to reaction time, error rate, and error type data. Error rate data alone may not offer enough information for the neuropsychologist interested in distinguishing between real improvements in performance and simple changes in strategy. For example, patients may slow their responses so that they can focus on accuracy instead of speed. Thus, results may incorrectly suggest an improvement in error rate across repeated attempts of a task. Other potential contributions for computerized testing have been detailed elsewhere (cf. Parsons, 2016).

In sum, computerized assessment offers both opportunities and challenges. The challenges can appear daunting but are a manageable and require informed consumers who can appreciate the issues and ask pertinent questions in evaluating their options. Evaluating computerized test systems also requires that test developers include detailed and specific information in their manuals specifying technical aspects, capabilities, and limitations of their systems with supporting data.

Disclosure statement

The authors report no conflicts of interest.

References


