Assessment
1–6

DOI: 10.1177/1073191114566622
asm.sagepub.com
SAGE

Validating the Accuracy of Reaction Time Assessment on Computer-Based Tablet Devices

Philip Schatz[1], Vincent Ybarra[1], and Donald Leitner[1]

Abstract

Computer-based assessment has evolved to tablet-based devices. Despite the availability of tablets and "apps," there is limited research validating their use. We documented 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, Samsung Galaxy) at response intervals of 100, 250, 500, and 1,000 milliseconds (ms). Results showed significantly greater timing error on Google Nexus and Samsung tablets (81–97 ms), than Kindle Fire and Apple iPads (27–33 ms). Within Apple devices, iOS 7 obtained significantly lower timing error than iOS 6. Simple reaction time (RT) trials (250 ms) on tablet devices represent 12% to 40% error (30–100 ms), depending on the device, which decreases considerably for choice RT trials (3–5% error at 1,000 ms). Results raise implications for using the same device for serial clinical assessment of RT using tablets, as well as the need for calibration of software and hardware.

Keywords

computer-based assessment, timing accuracy, reaction time

Following the advent of microcomputers, computer-based psychological assessment became a topic of interest, with an American Psychological Association (APA) Task Force offering guidelines in the 1980s (APA, 1986). Three decades later, the topic is still of interest within neuropsychology organizations, as evidenced by a *Joint Position Paper* of the American Academy of Clinical Neuropsychology (AACN) and the National Academy of Neuropsychology (NAN) on computerized neuropsychological assessment devices (Bauer et al., 2012).

The authors of the recent AACN and NAN *Position Paper* (Bauer et al., 2012) defined computerized neuropsychological assessment devices as ranging from "stand-alone computer-administered versions of established examiner-administered tests" to "fully web-integrated testing stations designed for general or specific applications," including "digital tablet, handheld device, or other digital interface" (p. 363). As of January 2014, approximately 42% of American adults (i.e., 18 or more years of age) owned a tablet computer (Pew, 2014). Applications for iPad and Android tablet devices are readily available for research purposes, as well as for the assessment of human neurocognitive abilities. Constructs such as auditory processing (Van Tasell & Folkeard, 2013), visual acuity (Black et al., 2013; Dorr, Lesmes, Lu, & Bex, 2013; Turpin, Lawson, & McKendrick, 2014), motor performance (Bertucco & Sanger, 2013), memory (Clionsky & Clionsky, 2014), and cognitive performance (Zhang, Red, Lin, Patel, & Sereno, 2013) have recently been assessed using tablet-based testing applications. Unfortunately, authors of only two of these studies validated the data acquired from a tablet device against traditional measures (Clionsky & Clionsky, 2014; Onoda et al., 2013). Others validated their tablet version against a computer-based version (Black et al., 2013; Dorr et al., 2013), or utilized only a tablet-based version with no criterion measure (Bertucco & Sanger, 2013; Van Tasell & Folkeard, 2013; Zhang et al., 2013). Those measuring reaction time (RT) or speed of responses have not yet established the validity of the tablet device for this purpose.

Personal computers (PCs) have been shown to have inherent limitations for measuring processing time, as the accuracy of time measurements is highly dependent on the computer type, speed, hardware, and software (McKinney, MacCormac, & Welsh-Bohmer, 1999). Others have identified the speed of the central processing unit (CPU) (MacInnes & Taylor, 2001), screen refresh rates, and peripheral input devices (Cernich, Brennana, Barker, &

[1]Department of Psychology, Saint Joseph's University, Philadelphia, PA, USA

Corresponding Author:
Philip Schatz, PhD, Professor of Psychology, Saint Joseph's University, 222 Post Hall, 5600 City Ave, Philadelphia PA 19131, USA.
Email: schatzSJU@gmail.com

Table 1. Tablet-Based Assessment Products.

Assessment product	Type of assessment
ImPACT	Concussion[ab]
C3 Logix	Concussion[ab]
Q-Interactive	WAIS-IV[ab], WISC-IV[ab], WMS-IV, NEPSY-II[ab]
Cambridge Cognition	ADHD[ab], Dementia[ab], Depression, Cognitive[a]
CAMCI	Cognitive[ab]

a. Indicates that the test uses simple reaction time.
b. Indicates that the test uses choice reaction time.

Bleiberg, 2007) as possible sources of timing errors. Millisecond (ms) timing accuracy has been achieved in both the Macintosh and Windows platforms (De Clercq, Crombez, Buysse, & Roeyers, 2003; MacInnes & Taylor, 2001; Westall, Perkey, & Chute, 1986, 1989); however, such accuracy often relies on a complicated combination of customized software and hardware.

Response time can be measured through simple and choice RT. Simple RT represents the time required to respond to a single stimulus, through a single response option, such as clicking a predetermined button (such as the "space bar" on the keyboard) when a stimulus appears (such as the screen changing color). Choice RT represents the time required to respond to one of two stimuli, each with its own response options—for example, clicking on the "A" button on the keyboard when a "left arrow" appears on the screen, versus clicking on the "L" button on the keyboard when a "right arrow" appears on the screen. Given the time required to respond to a single stimulus (e.g., simple RT), one can calculate decision time by subtracting simple RT from choice RT. As the stimuli become increasingly complex, and the time required to make a response increases, the resultant RT is more reflective of decision making than simple physiological ability.

Historically, measurement of RT in a clinical setting was contingent on mechanical apparatus (i.e., dating back to Wundt's "complication pendulum"), with average human (simple) RT documented as ranging from 150 (Seashore & Seashore, 1941) to 250 ms (Eckner, Kutcher, & Richardson, 2010). Among the benefits of computer-based testing are claims of ms timing accuracy, over traditional paper-based measures (Schatz & Zillmer, 2003). Currently, there is a growing trend in which traditional psychological tests are being modified for a touchscreen method of delivery (see Table 1). Most notably, the Wechsler Intelligence Scale for Children – 4th Edition (WISC–IV), Wechsler Adult Intelligence Scale – 4th Edition (WAIS–IV), California Verbal Learning Test – 2nd Edition and Children's Editions (CVLT–II, CVLT–C), and the Delis-Kaplan Executive Function System (D–KEFS) are available for iPad-based assessment, using "Q-interactive," which is a custom-developed iPad application that allows clinicians to administer clinical assessments via two tablets connected by Bluetooth technology (Pearson, 2013). Other custom assessment measures are being developed for administration using tablets devices, and all of which utilize some measure of simple RT. For instance, Cambridge Cognition's CANTAB attention-deficit/hyperactivity disorder (ADHD) Battery includes an RT test. The Computer Assessment of Mild Cognitive Impairment (CAMCI) is a tablet-based assessment of cognitive function in older individuals, incorporating RT measurement. C3 Logix is a concussion assessment app advertised by Apple Computers (Apple, 2014) that incorporates simple and choice RT among other cognitive measures.

Given that tablet devices are being already used to measure neurocognitive functioning, it can be expected that the number of available applications will increase. In this regard, the availability of the WAIS-IV in tablet format will likely increase use of tablet devices by clinicians, especially considering that the WAIS is the most frequently used assessment device among clinical psychologists (Camara, Nathan, & Puente, 2000).

From a clinical perspective, serial or longitudinal assessments require repeated administration of materials. Changes or variation in hardware or software used for computer-based assessment may introduce random error. The baseline/post-concussion assessment paradigm for clinical assessment of concussion demonstrates the analogous need for standardization and validation of such RT measures. Assuming an average RT "composite score" of 0.58 in a sample of high school athletes (Iverson, Brooks, Collins, & Lovell, 2006), a "reliable change" of 0.06 (or approximately a 10% slowing) represents significant decrease in performance beyond the standard error of measurement. If the accuracy of RT error varies by 5% to 10% between devices used, this could introduce error that represents the total amount of change required to document clinical decreases in performance, which would be erroneously attributed to the individual.

To date, the timing accuracy of tablet devices has not been established. The purpose of this study was to examine the accuracy of tablet devices within the context of the time span of simple and choice RT.

Method

Instruments

A custom external timing apparatus was created, a quartz-crystal-based precision time base module for which accuracy was validated within 5 ms with a Tektronix TDaS2024 oscilloscope. Stimulus presentation, in the form of a colored target that appeared on the screen, activated a phototransistor that, in turn, activated the timing apparatus. The timing apparatus then activated an electromechanical relay, which simulated a "touch response" via a standard electrocardiogram (ECG) conductive adhesive electrode secured to the screen of the tablet. The relay's closure, initiated by the timing apparatus, simulated a human touch by grounding the tablet's screen via

Table 2. Tablet Specifications.

Model	SoC (system on a chip)	Central processing unit	Graphics processing unit
iPad 3rd generation	Apple 5X	1 GHz dual-core ARM Cortex-A9	Quad-core PowerVR SGX543MP4
iPad 4th generation	Apple A6X	1.4 GHz dual-core AppleSwift	Quad-core PowerVR SGX554MP4
Kindle Fire 2nd generation	Texas Instruments OMAP	Dual-core 1.2 GHz TI OMAP4 4430	Imagination Technologies PowerVR
Google Nexus 7	Qualcomm Snapdragon S4 Pro	1.51 GHz quad-core Krait 300	400 MHz quad-core Adreno 320
Samsung Galaxy Tab 2	Cortex A9	TI OMAP4430 1.0 GHz dual-core	PowerVR SGX540

the electrode. The timing apparatus displayed the amount of time, in ms, between the appearance of the target on the tablet's screen and the activation of the relay. For the purpose of this study, we programmed the timing apparatus to delay the time between the appearance of the target on the tablet's screen and the activation of the relay (see Appendix A).

The following tablets were used, with the specified operating system (OS):

1) Apple iPad 3rd generation, iOS 6
2) Apple iPad 4th generation, iOS 6
3) Apple iPad 4th generation, iOS 7
4) Kindle Fire 2nd generation, Android 4.0
5) Google Nexus 7, Android 4.0
6) Samsung Galaxy Tablet 2, Android 4.0.

Tablet specifications can be found in Table 2. Each tablet was set to "airplane" mode, with no other apps running, and the brightness was level set at 70%.

The application used to test device RT was custom-developed for the purpose of this study. The application begins timing with the presentation of a visual stimulus on the screen, a standard blue box. The application ceases timing when a "touch signal" (i.e., finger down) interrupts the screen, with the delay between stimulus onset and screen touch presented in ms.

Validation of the timing device identified the following "inherent" delays:

- 10 ms from the initial appearance of the blue box on the tablet screen was required for the phototransistor's output to reach 5 volts DC and trigger the timing apparatus.
- 5 ms was required for the relay to close after its coil was energized.
- 20 ms of "contact" (i.e., relay closure grounding the electrode) on the device screen was required for the simulated touch to register.

Thus, the delay between the tablet's internal timing and the external timing apparatus was calculated by subtracting this 35 ms of accountable timing error.

Procedure

Each tablet was prepared for testing by being fully charged, closing all other applications, and setting the screen brightness. The tablet was then placed beneath the phototransistor and the electrode was attached to the screen. The app was activated, and after a uniform 2-second delay, the background color of the app changed, activating the phototransistor. Upon stimulus presentation, the activation of the phototransistor triggered the external timing apparatus, which, after a preprogrammed delay (i.e., a simulated RT) that was varied among 100 ms, 250 ms, 500 ms, and 1,000 ms, activated the "touch" relay and electrode on the tablet's screen. The timing delay (in ms) was then recorded, and the external timing device was reset. For each timing interval, 10 trials were performed on each tablet.

Analyses

A mixed-factorial-design multivariate analysis of variance (MANOVA) was used. The device/OS type served as the between-groups independent variable, and the RT interval served as the within-subjects independent variable. Mean timing delay between the internal device timer and the external timing device served as the dependent variable. Additional univariate analyses were conducted to identify between-device/OS differences within the Apple/iOS family as well as within the Android OS family. Partial eta-squared (η^2) was documented as a measure of effect size, with 0.01 constituting a small effect, 0.06 a medium effect, and 0.14 a large effect (Cohen, 1988).

Results

MANOVA revealed significant multivariate effects of device/OS [$F(5,216) = 808.5$, $p < .001$, $\eta^2 = .95$], interval [$F(3,216) = 10.5$, $p < .001$, $\eta^2 = .13$], and a significant device/OS by interval interaction [$F(15,216) = 4.5$, $p < .001$, $\eta^2 = .24$]. With respect to the main effect of device/OS, post-hoc Scheffé analyses identified significantly greater time delay on Samsung/Android (97 ms), which was greater than the Google Nexus/Android (81 ms), which was

Table 3. Main Effects of Device/OS and Interval on Timing Delay.

Device/OS	100 ms	250 ms	500 ms	1,000 ms
iPad 3 iOS 6	33.8 (4.5)	30.7 (6.2)	31.5 (7.4)	35.7 (7.7)
iPad 4 iOS 6	29.6 (15.9)	19.9 (1.1)	38.9 (0.9)	43.0 (5.0)
iPad 4 iOS 7[a]	27.9 (6.7)	27.2 (5.3)	27.9 (5.3)	29.8 (7.8)
Kindle Fire[b]	27.0 (3.0)	24.5 (2.9)	29.8 (5.4)	27.6 (3.5)
Google Nexus	85.6 (7.7)	72.5 (3.7)	83.5 (4.1)	85.1 (9.5)
Samsung Galaxy	98.3 (7.7)	99.3 (12.3)	98.0 (7.8)	94.0 (4.9)

Note. Numbers presented represent mean timing error, with standard deviation in parentheses.
a. iPad 4 iOS 7 < iPad 3 iOS 6, iPad 4 iOS 6.
b. Kindle Fire < Google Nexus < Samsung Galax.

greater than all three iPads and the Kindle Fire/Android (27–33 ms). With respect to the main effect of interval, post-hoc Scheffé analyses identified significantly lower time delay at the 250 ms interval (46 ms) than all the other intervals (50–57 ms). Means and standard deviations are provided in Table 3.

Within the Apple/iOS devices, 4th generation iPad running iOS 7 showed significantly lower timing error than either 3rd or 4th generation iPads running iOS 6 [$F(2,108) = 5.7$, $p = .004$, $\eta^2 = .10$]. Within the Android devices, Kindle Fire showed significantly lower timing error than did Google Nexus tablet, which showed significantly lower timing error than Samsung Galaxy [$F(2,108) = 1,211$, $p < .001$, $\eta^2 = .96$].

Discussion

This study is the first empirical validation of the timing accuracy of tablet-based devices. We identified significantly smaller timing delays on iPads (3rd generation running iOS6 or 4th generation running iOS 6 or 7) and Kindle Fire (running Android) than on both Google Nexus and Samsung Galaxy tablets (running Android). These differences were present at the subhuman RT interval of 100 ms, within the human simple RT interval of 250 ms, and within choice RT intervals of 500 and 1,000 ms.

We also documented the superiority of the iOS 7 operating system, within the Apple family of iPads, with small but significant effect size over iOS 6. Within the Android family, however, the Kindle Fire showed large and significant effect sizes over the Google Nexus and the Samsung Galaxy, and the Google Nexus small but significant effect size over the Samsung Galaxy.

Mean timing delay ranged from 27 to 98 ms, and screen refresh rates may account for 17 ms (60 Hz refresh rate; 1/60 second = 16.6 ms) of these delays. Overall, the unexplained timing delay in iPad/iOS devices and the Kindle Fire/Android device are operating within a margin of one to two screen refreshes. In contrast, the Samsung Galaxy and Google Nexus tablets running Android are operating well beyond the range of error that can be explained by screen refresh rates.

Putting these data into the context of clinical or research use, a 100 ms interval may have little utility as it is below the threshold of average human simple RT. However, unexplained timing error of 30 to 100 ms represents approximately 12% to 40% of a 250 ms timing interval, which is well within the range of human simple RT. As stimulus-response intervals increase from simple to choice RT, unexplained timing error of 30 to 100 ms represents only 6% to 10% of 500 ms responses and 3% to 5% of 1,000 ms responses. In this regard, the current findings have significant implications for the measurement of simple and choice RT on tablet-based devices. While the tablet/device/OS used may have little effect on the data when more complex choice RT trials are presented, inflated simple RT data can be expected depending on the tablet/device/OS used.

Despite these significant differences documented between tablet devices and timing intervals, the actual variation in timing accuracy (as measured by standard deviation) was quite consistent. This suggests that while specific tablet devices have greater inherent error, this error is standardized and consistent. Software developers should be aware of these differences and should either provide parameters for interpretation of reaction time data or conduct internal calibration or normalization of data prior to clinical output.

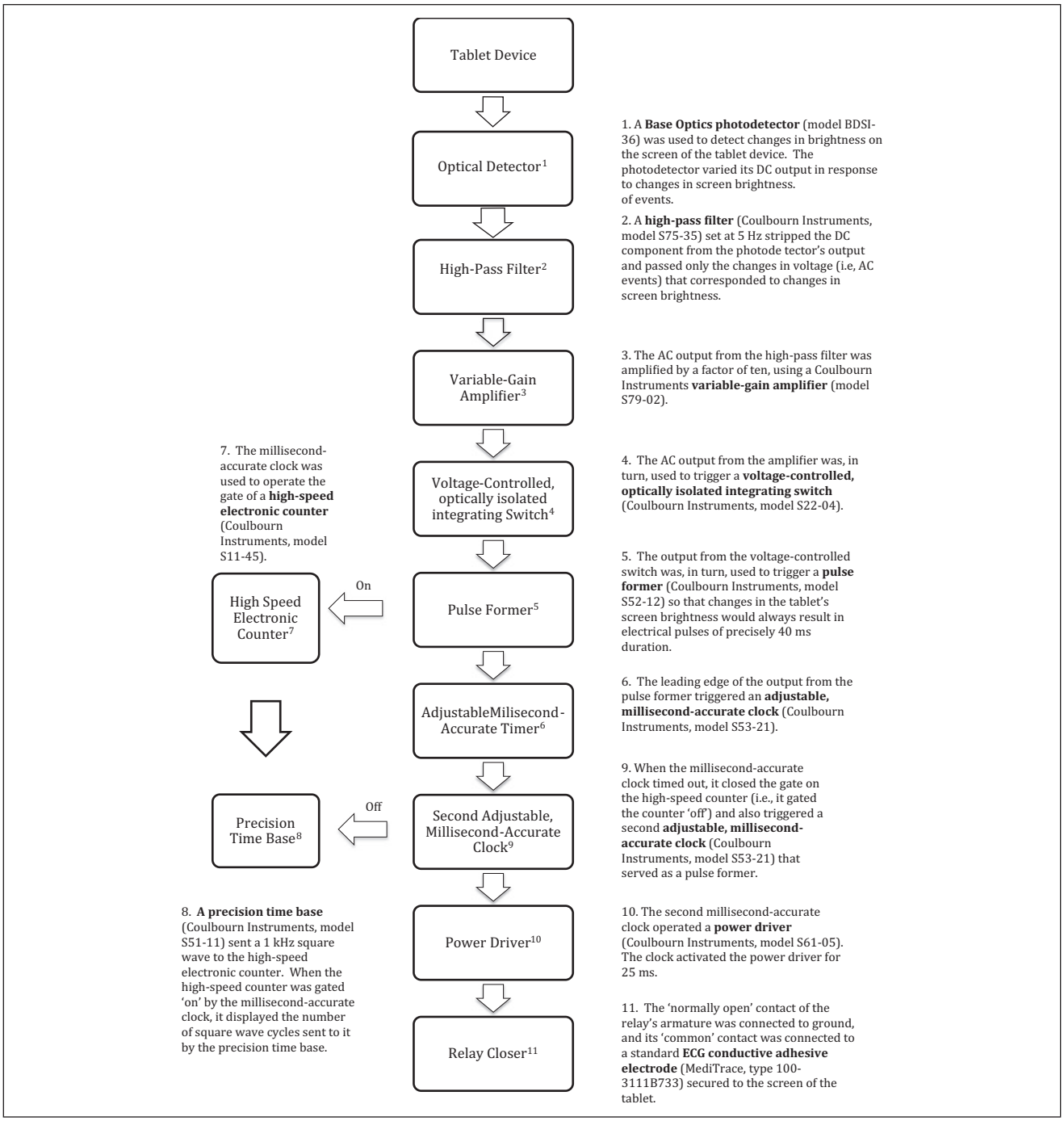
The present study is not without its limitations. While every step was taken to replicate the touch response by a human on a tablet device, the use of an electrified conductive electrode may not duplicate the exact nature of a human finger. In addition, while variation in timing error was quite small, and consistent from trial to trial, we ran a limited number of trials at each response interval; with increased trials, variation might decrease. As well, while we used basic simple and choice RT tasks, these tasks were not extracted from or intended to mirror those from commercial products. While we expect the same results would be obtained from these products, the current results may be esoteric to the tasks employed. Finally, while we attempted to utilize the most modern tablet devices, running the most current operating systems, technological advances often

occur faster than scholarly empirical research. As newer devices and operating systems become available, replication of these results may be warranted.

Clinicians using, or considering switching to, tablet-based assessments should be aware of the need for consistent use of devices, especially when administering serial assessments. More explicitly, results from paper-based or computer-based assessment should not be considered equivalent or comparable. Even when results from tablet-based assessments are available for comparison, clinicians should be aware of the device and operating system used, especially when considering RT data.

Appendix A

Timing Apparatus



Declaration of Conflicting Interests

Dr. Schatz has served as a consultant to ImPACT Applications Inc., although ImPACT had no role in the conceptualization of the study, the collection or analysis of data, the writing of the manuscript, or the decision to submit it for publication. Mr. Ybarra and Dr. Leitner have no conflict of interest to declare.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

American Psychological Association (APA). (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: Author.

Apple. (2014). *A new game plan for concussions*. Retrieved November 5, 2014, from https://http://www.apple.com/your-verse/concussion-game-plan/

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: Joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Arch Clin Neuropsychol*, *27*(3), 362–373.

Bertucco, M., & Sanger, T. D. (2013). Speed-accuracy testing on the Apple iPad(R) provides a quantitative test of upper extremity motor performance in children with dystonia. *J Child Neurol*. doi:10.1177/0883073813494265

Black, J. M., Jacobs, R. J., Phillips, G., Chen, L., Tan, E., Tran, A., & Thompson, B. (2013). An assessment of the iPad as a testing platform for distance visual acuity in adults. *BMJ Open*, *3*(6). doi:10.1136/bmjopen-2013-002730

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implication in professional psychology. *Professional Psychology: Research and Practice*, *31*(2), 141–154.

Cernich, A. N., Brennana, D. M., Barker, L. M., & Bleiberg, J. (2007). Sources of error in computerized neuropsychological assessment. *Arch Clin Neuropsychol*, *22*(Suppl. 1), S39–48.

Clionsky, M., & Clionsky, E. (2014). Psychometric equivalence of a paper-based and computerized (iPad) version of the Memory Orientation Screening Test (MOST). *Clin Neuropsychol*, *28*, 1–9. doi:10.1080/13854046.2014.913686

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). San Diego, CA: Academic Press.

De Clercq, A., Crombez, G., Buysse, A., & Roeyers, H. (2003). A simple and sensitive method to measure timing accuracy. *Behav Res Methods Instrum Comput*, *35*(1), 109–115.

Dorr, M., Lesmes, L. A., Lu, Z. L., & Bex, P. J. (2013). Rapid and reliable assessment of the contrast sensitivity function on an iPad. *Invest Ophthalmol Vis Sci*, *54*(12), 7266–7273. doi:10.1167/iovs.13-11743

Eckner, J. T., Kutcher, J. S., & Richardson, J. K. (2010). Pilot evaluation of a novel clinical test of reaction time in National Collegiate Athletic Association Division I football players. *J Athl Train*, *45*(4), 327–332. doi:10.4085/1062-6050-45.4.327

Iverson, G. L., Brooks, B. L., Collins, M. W., & Lovell, M. R. (2006). Tracking neuropsychological recovery following concussion in sport. *Brain Inj*, *20*(3), 245–252.

MacInnes, W. J., & Taylor, T. L. (2001). Millisecond timing on PCs and Macs. *Behav Res Methods Instrum Comput*, *33*(2), 174–178.

McKinney, C. J., MacCormac, E. R., & Welsh-Bohmer, K. A. (1999). Hardware and software for tachistoscopy: How to make accurate measurements on any PC utilizing the Microsoft Windows operating system. *Behav Res Methods Instrum Comput*, *31*(1), 129–136.

Onoda, K., Hamano, T., Nabika, Y., Aoyama, A., Takayoshi, H., Nakagawa, T., . . . Yamaguchi, S. (2013). Validation of a new mass screening tool for cognitive impairment: Cognitive Assessment for Dementia, iPad version. *Clin Interv Aging*, *8*, 353–360. doi:10.2147/CIA.S42342

Pearson. (2013). *Q-interactive, groundbreaking mobile solution for interactive assessments, now available to qualified clinicians*. Retrieved November 5, 2014, from http://www.pearsoned.com/q-interactive-groundbreaking-mobile-solution-for-interactive-assessments-now-available-to-qualified-clinicians/

Pew. (2014). *Internet Project Omnibus Survey*. Washington, DC: Author.

Schatz, P., & Zillmer, E. A. (2003). Computer-based assessment of sports-related concussion. *Appl Neuropsychol*, *10*(1), 42–47.

Seashore, S. H., & Seashore, R. H. (1941). Individual differences in simple auditory reaction times of hands, feet and jaws. *Journal of Experimental Psychology*, *29*(4), 342–345.

Turpin, A., Lawson, D. J., & McKendrick, A. M. (2014). PsyPad: A platform for visual psychophysics on the iPad. *J Vis*, *14*(3), 16. doi: 10.1167/14.3.16

Van Tasell, D. J., & Folkeard, P. (2013). Reliability and accuracy of a method of adjustment for self-measurement of auditory thresholds. *Otol Neurotol*, *34*(1), 9–15. doi:10.1097/MAO.0b013e318278c05d

Westall, R., Perkey, M. N., & Chute, D. L. (1986). Accurate millisecond timing on Apple's Macintosh using Drexel's Millitimer. *Behavior Research Methods Instruments and Computers*, *18*, 307–311.

Westall, R., Perkey, M. N., & Chute, D. L. (1989). Millisecond timing on Apple's Macintosh revisited. *Behavior Research Methods Instruments and Computers*, *21*, 540–547.

Zhang, M. R., Red, S. D., Lin, A. H., Patel, S. S., & Sereno, A. B. (2013). Evidence of cognitive dysfunction after soccer playing with ball heading using a novel tablet-based approach. *PLoS One*, *8*(2), e57364. doi:10.1371/journal.pone.0057364