The psychology experiment building language (PEBL) and PEBL test battery

Article *in* Journal of Neuroscience Methods · November 2013

DOI: 10.1016/j.jneumeth.2013.10.024 · Source: PubMed

2 authors:

Shane T. Mueller
Michigan Technological University

Brian Piper
Bowdoin College


Volume 222, 30 January 2014
ISSN 0165-0270
ELSEVIER
JOURNAL OF
NEUROSCIENCE
METHODS
Editor-in-Chief
Vincenzo Crunelli

Journal of Neuroscience Methods 222 (2014) 250–259

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Basic Neuroscience

The Psychology Experiment Building Language (PEBL) and PEBL Test Battery

Shane T. Mueller[a,*], Brian J. Piper[b]

[a] Department of Cognitive and Learning Sciences, Michigan Technological University, United States
[b] Department of Basic Pharmaceutical Sciences, Husson University, United States

ARTICLE INFO

Article history:
Received 8 August 2013
Received in revised form 24 October 2013
Accepted 30 October 2013

Keywords:
Open science
Iowa Gambling Task
Wisconsin Card Sorting Test
Continuous Performance Test
Trail-making Test
Executive function
Mental rotation
Motor learning

ABSTRACT

Background: We briefly describe the Psychology Experiment Building Language (PEBL), an open source software system for designing and running psychological experiments.
New method: We describe the PEBL Test Battery, a set of approximately 70 behavioral tests which can be freely used, shared, and modified. Included is a comprehensive set of past research upon which tests in the battery are based.
Results: We report the results of benchmark tests that establish the timing precision of PEBL.
Comparison with existing method: We consider alternatives to the PEBL system and battery tests.
Conclusions: We conclude with a discussion of the ethical factors involved in the open source testing movement.

1. Introduction

The Psychology Experiment Building Language (PEBL) is a free, open-source software system that allows researchers and clinicians to design, run, and share behavioral tests. At its core, PEBL is a programming language and interpreter/compiler designed to make experiment writing easy. It is cross-platform, written in C++, and relies on a Flex/Bison parser to interpret programming code that controls stimulus presentation, response collection, and data recording. PEBL is designed to be an open system, and is licensed under the GNU Public License 2.0. This allows users to freely install the software on as many computers as they wish, to share their experiments with others without worrying about licenses, to distribute working experiments to other researchers or remote subjects without requiring special hardware locks, and to examine and improve the system itself when it does not suit one's needs.

2. History of the Psychology Experiment Building Language

Development and design of PEBL began in 2002. An initial limited release of PEBL 0.1 was made in 2003, with the first public release to the Sourceforge.net servers in January 2004. The initial motivation for its development and design was a dissatisfaction with the current experiment-running systems available. At the time, there were no robust cross-platform Free systems available, which meant that major vendors would focus on one platform (e.g., Psyscope for Macintosh computers, Superlab or E-Prime for Windows PCs, and no packages available for Linux), or researchers who cared about cross-platform testing would develop one-off cross-platform applications using Java or web interfaces. The original design of PEBL abstracted aspects of stimuli, response-collection, and data structures so that it could be implemented on multiple distinct platforms. However, the first implementation platform was done via the Simple DirectMedia Layer (http://libsdl.org) gaming library, which itself is a cross-platform library. Because of this, versions of PEBL are available on MS Windows, OSX, and Linux operating systems.

Due to its limited capabilities, initial use and adoption of PEBL was fairly modest. For the first year, roughly 250 users downloaded PEBL, and five emails were exchanged on the support email list. During 2005, PEBL's activity improved somewhat, nearly doubling to 450 downloads and 10 emails exchanged. Over this period, five versions of PEBL were released (0.1–0.5). Starting in 2006, we began to release an accompanying test battery, initially consisting of eight commonly-used laboratory tests, which is primarily responsible for the first increase in downloads, and an initial set of publications starting in 2008. Since then, the number of downloads have increased to a stable level of between 1000 and 2000 downloads

* Corresponding author at: Department of Cognitive and Learning Sciences, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, United States. Tel.: +1 906 487 1159.
E-mail address: shanem@mtu.edu (S.T. Mueller).

0165-0270/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jneumeth.2013.10.024

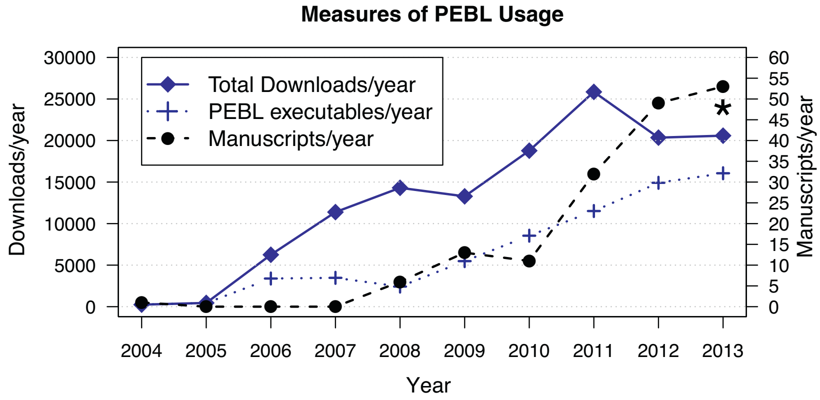

Fig. 1. Three measures of the usage and adoption of PEBL over time: the number of downloads (total and PEBL installer only) recorded via sourceforge.net, and the number of published manuscripts (papers, conference proceedings, theses, etc.). *2013 figures are partial values through November 30, 2013.

per month (see Fig. 1, which also shows the number of published manuscripts that have used or cited PEBL over time[1].) For 2012, the last year for which we have complete records, there were 50 publications citing PEBL, and 20,300 downloads.

The true usage of PEBL is likely to be much broader than may be reflected by citation counts for several reasons. First, PEBL gets used frequently in research methods and other undergraduate courses, resulting in studies that cannot typically be published. Second, PEBL has been used in many academic theses (including bachelor's-level honors theses up to Ph.D theses), and for research studies that are only presented at conferences, and these are not systematically indexed or publicized. Third, because it is free, PEBL gets used internationally, where research studies may be more likely to be presented only at regional conferences or in language-specific journals that are not indexed. Finally, even the top journals systematically fail to require authors to cite work related to the software they use to conduct the study. Many publications that use PEBL have merely referenced its website in a footnote or parenthetical comment. These footnoted references do not appear in standard citation indexes, and it is likely that there are published articles that have used PEBL but made no reference or mention to the system by name.

3. Features of PEBL

The features of PEBL are far too numerous to describe in detail here. We provide a 200-page reference manual that can be freely downloaded or purchased on-line which details the programming system and its functions (Mueller, 2012a,b). As a basic overview, PEBL 0.13 supports a number of stimulus types, including images (in a variety of image formats), text rendered in TrueType fonts using both single-line stimuli and multi-line text objects; many rendered shape primitives (lines, circles, rectangles, etc.); audio recordings, video recordings, and simple generated sounds. For response collection, PEBL supports keyboard, mouse, gaming device input, communication via TCP/IP, serial, and parallel port, and a software audio voice key. In addition, timing of stimuli and responses can be recorded and controlled with a precision dictated primarily by the limits of the hardware and operating system used. As we will show later in the manuscript, internal event timing can be scheduled/performed with 1-ms precision, keyboard responses can be recorded with roughly 5-ms precision, and stimuli can be displayed in increments of the video update frequency.

[1] A complete listing is available at http://pebl.sourceforge.net/wiki/index.php/Publications_citing_PEBL.

PEBL provides a library of functions for general computing as well as ones devoted to the design of experiments. These include a wide selection of functions for randomization, sampling, and counterbalancing; data handling and statistics; standard experimental idioms (e.g., built-in functions for messages, multiple-choice questions, many types of commonly-used visual stimuli), and both restricted-set (e.g., press one of several keyboard buttons) and multidimensional response collection (e.g., free-form typed input).

PEBL experiments are typically run via a software launcher that allows users to select aspects of how the test is conducted (screen resolution, participant code, etc.) and also allows "experiment chains"; tests that can be run in sequence. The launcher is itself written in PEBL, and so achieves cross-platform execution on any platform PEBL is available on. A screenshot of the PEBL launcher is shown in Fig. 2.

In comparison to other similar systems, PEBL has both advantages and weaknesses. For example, a number of similar systems employ a special-purpose GUI that can be used for visual programming to create simple experiments using drag-and-drop metaphors, including OpenSesame (Mathôt et al., 2012); PsyScopeX (http://psy.cns.sissa.it/) E-Prime (http://pstnet.com), Presentation (http://neurobs.com), and others. In contrast, PEBL experiments are implemented via a flexible, full-featured programming language, which limits is accessibility to some users, although enabling more elaborate experimental designs. Detailed instructions and tutorials for programming in PEBL are available elsewhere, but we have included the code for a simple choice-response experiment as an Appendix. In addition, PEBL is designed to avoid many object-oriented and event-focused programming metaphors found in full-featured GUI programming toolkits which are sometimes confusing to novices.

For use as a scientific tool, the open-source nature of PEBL has advantages over many closed-source solutions. These advantages begin with the ability to inspect, alter, and redistribute the source code, so that experimenters can verify and change aspects of an experiment, and an experimentation tool can live on even if the original developer abandons it. Another advantage is that the development model enables using a large number of existing open source libraries and source code developed by others, reducing the complexity of PEBL. Finally, the open-source nature means that executables can be freely distributed, allowing experimenters great flexibility in how they conduct their tests. Nevertheless, it should be recognized that commercial software generates revenues that can help to support norming studies, continued software development, documentation, bug fixing, and can provide dedicated support to customers.

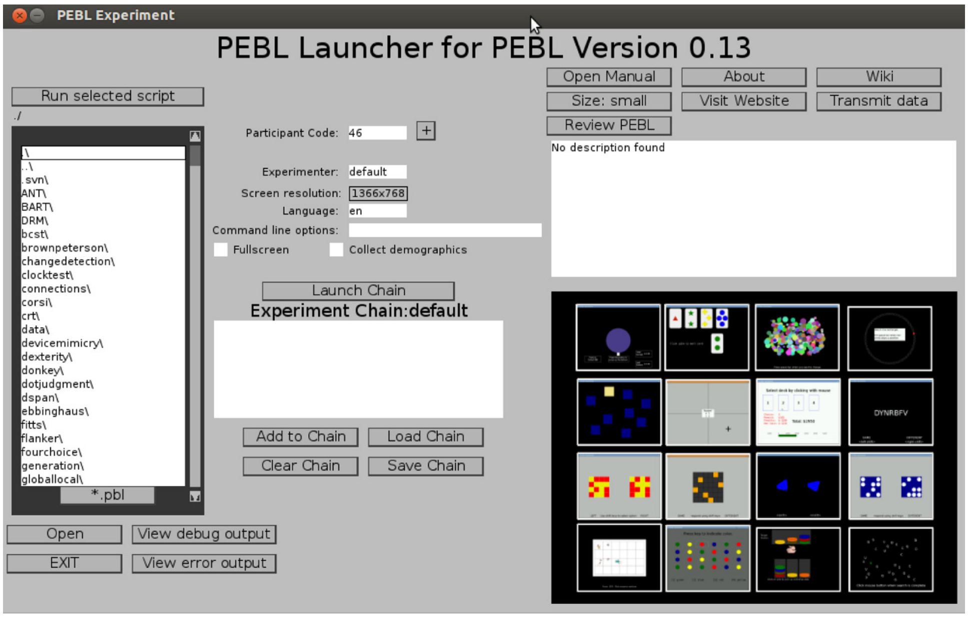

Fig. 2. Screenshot of the PEBL Launcher, which allows a user to navigate the test battery, run specific tests, and execute a 'chain' of tests appropriate for a particular study.

PEBL itself is compiled code written in C++, which means users wishing to modify PEBL itself must have the ability to code in C++ and recompile the system. A number of alternate existing systems are based on the popular open-source interpreted Python language (see Geller et al., 2007; Peirce, 2007; Straw, 2008; Kötter, 2009; Krause and Lindemann, 2013), which may have a lower barrier to entry for developers. However, PEBL is distributed as an all-in-one system, and does not require downloading and installing separate software libraries, making it more accessible to non-technical users than many of these approaches. PEBL also comes with perhaps the largest available open-source test battery in existence, which means that many users never need to create their own new tests, but rather use or modify existing ones.

4. The PEBL Test Battery

First released in mid-2006, the objective of the PEBL Test Battery is to bring open-source versions of common testing paradigms to the research and clinical testing community. Each subsequent release of the main PEBL system has been accompanied by an updated test battery which includes improvements, bug fixes, and new tests. Each test attempts to provide a reasonable stand-alone implementation of the paradigm, but typically offers the ability to change and alter the task to suit the needs of a particular study. The test battery focuses on computerized cognitive tests rather than questionnaire-based personality tests, including paradigms that involve memory, attention, and executive control. Over the past six years, the test battery has grown to roughly 70–80 tests and test variants, which are described in Table 1. The more popular tests have been translated by users into many languages, and these tests have been adapted so that the chosen language can be selected via the launcher, falling back on English if no translation is available.

Starting around 2008, the first wave of experiments using PEBL began to be published (see Fig. 1). Since that time, we have identified more than 150 published articles, theses, reports, conference presentations, and similar publications that have used or cited PEBL. These cover disciplines including Artificial Intelligence (Mueller, 2010), cognitive psychology (Worthy et al., 2013), neurology (Clark and Kar, 2011; Kalinowska-Łyszczarz et al., 2012), clinical psychology (Gullo and Stieger, 2011), cognitive neuroscience (Danckert et al., 2012), behavioral endocrinology (Premkumar et al., 2013), medical education (Aggarwal et al., 2011), neuropharmacology (Lyvers and Tobias-Webb, 2010), physiology (Piquet et al., 2011), developmental neuroscience (Piper, 2011), genetics (Ness et al., 2011), computer science (Cinaz et al., 2012), and human factors (Lipnicki et al., 2009; Qiu and Helbig, 2012).

The PEBL Launcher will allow a user to navigate the test battery directly. As the directory tree containing the battery is navigated, a screenshot of the particular test will appear on the screen, along with basic information regarding the test. When a test is selected, clicking the 'wiki' button will open a web page that will provide more detailed information about the selected test, including information about data file formats, background references, options, and the like.

Although each test is different and saves different information to data files, there are some basic similarities across tests. Data files are saved in a data\ sub-directory within the test's directory, within the pebl-exp-0.13 location. Most tests save a basic data file in .csv format that writes a single line of data per trial or response, with a separate file saved for each participant (and the filename incorporating the participant code). Each line of the file includes both independent variables (subject code, trial number, block codes, condition codes, etc.) and dependent measures (response, response time, accuracy, etc.). These files are typically saved with a filename identifying the participant code, although as a safety mechanism filenames that already exist will not be overwritten, but rather a new related filename will be created by appending a number to the end of the file's base name. Typically, all of the data files for an experiment can be concatenated together to form a single master file (usually removing the first row, which often includes column headers) for analysis using statistical software.

In addition to the detailed data file, several other files are sometimes created. For many tests, a human readable report file is created that will record mean accuracy and response times across important conditions of the study. Also, some scripts create summary files recording a single row of independent and dependent variables per participant, as well as pooled data files that combine all detailed data into a single .csv file.

Table 1
Description of the roughly 70 tasks and task variants available as part of the PEBL Test Battery.

Test name	Directory name	Description	References
Aimed Movement Test	fitts	Use mouse to point to targets of different sizes and distances	Fitts (1954), Mueller (2010)[a]
Attentional Network Test	ANT	Combines orienting, cuing, and flanker-based filtered attention	Fan et al. (2002)
Balloon Analog Risk Task	BART	Inflate a balloon for rewards	Lejuez et al. (2002)
Bechara's "Iowa" Gambling Task	iowa	Choose decks with different rewards and penalties	Bechara et al. (1994)
Berg's "Wisconsin" Card Sorting Test	bcst	Executive function feature-switching task (normal and 64-card version)	Berg (1948)
Brown-Peterson Task	brownpeterson	Test of short-term memory decay	Peterson and Peterson (1959)
Choice Response Time Task	crt	Choose between several options	e.g., Logan et al. (1984)
Compensatory Tracking Task	ptracker	Move dynamic cursor to target	Ahonen et al. (2012)[a]
Connections Test	connections	Version of trail-making task with minimal motor requirements	Salthouse et al. (2000)
Continuous Performance Task	pcpt	Respond to a visual stimulus	Conners et al. (2003), Piper (2012)[a]
Corsi Block-tapping Test	corsi	Short-term memory for visual sequences (Forward and Backward)	Corsi (1972), Kessels et al. (2000, 2008)
Device Mimicry	devicemimicry	Control articulated 8-DF device to reproduce trail	Mueller (2010)[a]
Digit Span	dspan	Simple digit span task	Croschere et al. (2012)[a]
Dot Judgment Task	dotjudgment	Determine which field has more dots	Damos and Gibb (1986)
Dual N-Back	nback	Dynamic working memory task; remember two memory streams	Jaeggi et al. (2008)
Ebbinghaus Memory Test	ebbinghaus	Learn and recall sequences of nonsense syllables	Ebbinghaus (1913)
Flanker Test	flanker	Respond to center stimulus in background of incongruent flankers	Eriksen and Schultz (1979), Stins et al. (2007)
Four-choice Response Task	fourchoice	Choose one of four responses	Wilkinson and Houghton (1975), Perez et al. (1987)
Generation Effect	generation	Determine role of generation on recall	Hirshman and Bjork (1986)
Go/No-Go task	gonogo	Respond to one stimulus; ignore second stimulus	Bezdjian et al. (2009)
Hungry Donkey Test	donkey	Children's analog of Bechara's Iowa Gambling Task	Crone and van der Molen (2004)
Implicit Association Task	iat	Combine two parallel decision processes to assess implicit associations	Greenwald et al. (1998)
Item/Order Task	itemorder	Remember either order or content of letter strings	Perez et al. (1987)
Letter-Digit Substitution Task	letterdigit	Learn mapping between letters and numbers	Perez et al. (1987)
Lexical Decision Task	lexicaldecision	Determine whether a stimulus is a ward	Meyer and Schvaneveldt (1971)
Mackworth Clock Test	clocktest	Sustained attention task; watch clock that skips a beat	Mackworth (1948)
Manikin Task	manikin	Three-dimensional rotation task	Carter and Woldstad (1985)
Match-to-sample Task	matchtosample	Remember visual pattern and compare to test pattern	Perez et al. (1987), Ahonen et al. (2012)[a]
Math Processing	mathproc	Simple math task	Perez et al. (1987)
Math Test	mathtest	Tests solving math problems (2)	Novel Tasks
Matrix Rotation	matrixrotation	Mentally rotate a visual grid	Phillips (1974), Perez et al. (1987)
Memory Scanning	sternberg	Sternberg's classic memory scanning paradigm	Sternberg (1966)
Memory Span	mspan	Memory span task with spatial response mode	Croschere et al. (2012)[a]
Mental Rotation Task	rotation	Rotate polygon to determine whether it matches target.	Berteau-Pavy et al. (2011)[a]
Mouse Dexterity Test	dexterity	Move noisy mouse to a target	Novel Task
Mueller-Lyer Illusion	mullerlyer	Psychometrically determine size of illusion using staircase method	Müller-Lyer (1889)
Object Judgment Task	objectjudgment	Determine whether Attneave shapes are same or different in size, shape (2 tasks)	Mueller (2010)[a]
Oddball Task	oddball	Respond to low-probability dimension	Huettel and McCarthy (2004)
Partial Report Task	partial-report	Encode visual array and respond with identity of cued subset	Lu et al. (2005)
PEBL Card-sorting Test	pcards	Dimensional binary card-sorting test with 10 dimensions	Novel Task
PEBL Switcher Task	switcher	Select target matching a specific dimension of previous target	Anderson et al. (2012)[a]
Posner Cueing Task	spatialcuing	Respond to stimulus in probabalistically-cued location	Posner (1980), Lu et al. (2005)
Probabilistic Reversal Learning	probrev	Learn when a probabilistic rule shifts	Cools et al. (2002)
Probability Monitor	probmonitor	Detect when a dynamic gauge shows a signal	Perez et al. (1987)
Probe Digit Task	probedigit	Short-term recall of number sequence elements	Waugh and Norman (1965)
Psychomotor Vigilance Task	ppvt	Vigilance task used to measure sleepiness	Dinges and Powell (1985), Ulrich (2012)[a]
Pursuit Rotor Task	pursuitrotor	Move cursor to follow smoothly moving target	Piper (2011)[a]
Random Number Generation	randomgeneration	Generate sequence of random numbers; Test of executive function.	Towse and Neil (1998)
Ratings Scales	scales	Several subjective ratings scales, including NASA-TLX, tiredness, heat/comfort	Hart and Staveland (1988)
Rhythmic Tapping Task	timetap	Tap at a prescribed rate for 3 minutes	Perez et al. (1987) (Test 19)
Simon Task	simon	Stimulus-response compatibility test	Yamaguchi and Proctor (2012)
Simple Response Time	srt	Make response to onset of a stimulus	e.g., Logan et al. (1984)
Situation Awareness Task	satest	Dynamic visual tracking task	Tumkaya et al. (2013)[a]
Spatial Priming	spatial-priming	Respond to a stimulus in a 3 × 3 grid with location primes	Nelson (2013)[a]
Speeded tapping/Oscillation Task	tapping	Tap as fast as possible for 60 s	Freeman (1940)
Stroop interference Task(s)	stroop	Make responses to one dimension of multi-dimensional stimulus (color stroop x2, number stroop, victoria stroop)	Troyer et al. (2006)
Survey Generator	survey	Create simple survey using a spreadsheet	Novel task
Symbol-counter Task	symbolcounter	Keep track of matching symbols	Garavan (1998), Gehring et al. (2003)
Test of Attentional Vigilance	toav	Respond to one kind of non-verbal stimulus; implementation of TOVA	Greenberg et al. (1996), Ahonen et al. (2012)[a]
Time Wall	timewall	Judge when an occluded moving target will reach destination	Jerison et al. (1957), Piper et al. (2012)[a]
Tower of Hanoi	toh	Solve classic tower puzzle	Kotovsky et al. (1985)
Tower of London	tol	Solve tower puzzle used to understand planning (10 versions)	Anderson et al. (2012)[a]
Trail-making Test	ptrails	Connect-the-dots task requiring executive switching	Piper et al. (2012)[a]
Two-column Addition	twocoladd	Add three 2-digit numbers	Perez et al. (1987)
Typing task	typing	Type various passages of text	Novel task
Visual Change Detection	changedetection	Look for changes in color, size, location of random circles	Novel task
Visual Pattern Comparison	patterncomparison	Compare two pattern grids (3 versions)	Perez et al. (1987)
Visual Search	vsearch	Search for target in distractors	Treisman (1985)

[a] References include both sources where original tasks were published upon which the PEBL version is based, and recent publications using the actual PEBL tasks.

5. Citing PEBL

Different citation formats have different practices for citing software. The American Psychological Association (APA) guidelines suggest that software such as PEBL can be cited in a format such as:

Mueller, S. T. (2013). The Psychology Experiment Building Language (Version 0.13) [Software]. Available from http://pebl.sourceforge.net

6. Evaluation of timing properties of PEBL

For many neuropsychological tests, researchers are interested in the time measurement precision of the software, and this is true for many tests in the PEBL Test Battery, where precise timing is an important aspect of the psychological test. Timing errors can impact three primary aspects of psychological testing: the timing of events within a test, the precise timing of visual stimulus onsets and offsets, and the precise timing of responses. In each of these cases, the timing of events may need to be logged and registered against a highly-precise clock. In the following section, we developed three testing/benchmarking scripts to assess the timing properties of the PEBL. These scripts are available by request from the authors, and will be included in future versions of PEBL.

PEBL uses a computer's real-time clock (RTC) to measure current time, in ms, from the point at which the particular experiment began. This level of precision is typically sufficient for most psychological and neuropsychological testing, and this clock forms the basis for all timing functions within PEBL. In PEBL, two functions form the basis of most timing: the `GetTime()` function (which returns the clock time) and the `Wait()` function (which delays a specific number of milliseconds).

6.1. Wait timing and clock access

The `Wait()` function takes a delay (in ms) as an argument, schedules a particular test to be evaluated within PEBL's event loop, which runs repeatedly until the test is satisfied. The test scheduled by `Wait()` will be satisfied once the RTC value is greater than the delay plus the value of the RTC when the event `Wait()` function began.

The event loop can run in two modes, depending on the value of a global variable called `gSleepEasy`. If `gSleepEasy` is non-zero, the PEBL process is put to 'sleep' for a short period at the end of each execution of the event loop, waking up through the use of an interrupt that will occur at earliest one computer interrupt step later. This sleep gives the computer a chance to catch up with other pending processes, and can sometimes improve overall timing precision. However, depending on the hardware, operating system, and particular settings, this time can be as long as 10 ms, meaning that if an event occurs during that sleep, it will not be recorded until at earliest when the interrupt is handled again. If another process has a higher priority, the operating system may not return to the dormant process for several steps, reducing time precision further.

If the variable `gSleepEasy` is 0, the process is not put into sleep mode during the event loop, creating a 'busy wait' where the RTC may be tested many times every ms, reducing the chance of the process being delayed. Although one might assume that this will give better timing precision (and at times it does), it may not always do so, because an OS may identify the process as being too greedy and reduce its priority.

To understand how PEBL performs using `Wait()` commands in these two scenarios, we developed a PEBL script that tested the observed timing of random `Wait()` commands. All testing reported here was conducted on a Dell Precision T1600 PC running Windows 7, using a Planar PX2230MW monitor at a resolution of 1920 × 1080. In the current study, on 1000 consecutive trials, a random number between 1 and 200 was sampled, and a `Wait()` command was issued with that argument. Immediately before and after the command, the RTC clock time was recorded using the `GetTime()` function. Then, the actual time of the wait was recorded along with the programmed time. This was done under both 'easy' and 'busy' wait settings.

In the 'busy' wait condition, every trial (1000/1000) was measured to take exactly the same duration as the programmed time. In contrast, for the 'easy' wait condition, no trials (0/1000) were identical to the programmed time, but 941/1000 were 1 ms longer than the programmed time and the remaining 59 were 2 ms longer. The correlation between time over and programmed time was not significant ($R = .035$, $t(998) = 1.1$, $p = .26$), indicating that there was no relationship between how long the wait was and the size of the overage. Although the difference between busy and easy waits is clearly statistically reliable, is likely of little consequence for most applications, as the overage was always less than 2 ms. However, the sleep setting may have greater effects for experiments that are more complex, run on systems that have other concurrent processes, less computational resources, or possibly those that access hardware input or output.

6.2. Response timing with a gaming keyboard

Timing precision of responses can be impacted by timing issues that exist in the `Wait()` function, as well as other factors related to detecting and processing keypresses. To assess the precision of response timing, we developed a second PEBL script that records the timing of a keypress for five 20-s trials. We then adapted a Lafayette Instruments Illusionator Model 14014 device, which is essentially a motor whose rotation speed can be controlled via a dial. We secured a standard compact disk off-center on the rotation axis of the device, to act as a cam that could depress a keyboard key once per rotation. Tests were performed using the keypad 'Enter' key on a Razer BlackWidow gaming keyboard. The BlackWidow uses high-precision mechanical keyboard switches and special internal circuitry that putatively allows the keyboard state to be polled 1000 times per second (in contrast to most commercial keyboards, whose polling frequency may be much lower, and whose rate is typically undocumented). We selected three basic inter-press durations; roughly 100, 200, and 300 ms/press. As a reference, in testing, the fastest we were able to press the key with a finger was with an inter-response time of about 170 ms.

Once the responses were recorded, we computed the time between consecutive key-presses for each condition, to establish the extent to which response times were systematically recorded. Fig. 3 shows histograms of these experiments, with the left panels showing absolute inter-response times, and the right panels showing the times relative to the mean time for each 20-s block (this was done to debias potential drifts in the rotation rate across trials). In each of the panels, it is apparent that recorded inter-response times are clustered at roughly 5-ms epochs. Fig. 4 depicts a particular set of 100 responses from the fast tapping condition, which shows that the recorded inter-response time tended to oscillate between 95 and 100 ms, with occasional differences of 1 ms. Although this oscillation could in theory stem from a second-order oscillator in the physical device used to depress the key, we believe it is more likely due to the operating system, gaming library, or keyboard's rate of polling devices and handling input events, which appears to have a quantum unit of 5 ms. Overall, debiased response distributions (right column) are mostly within a 5 ms time-frame of the average, indicating that this is a limiting timing precision on the system we tested. Other tests using an off-the-shelf keyboard have

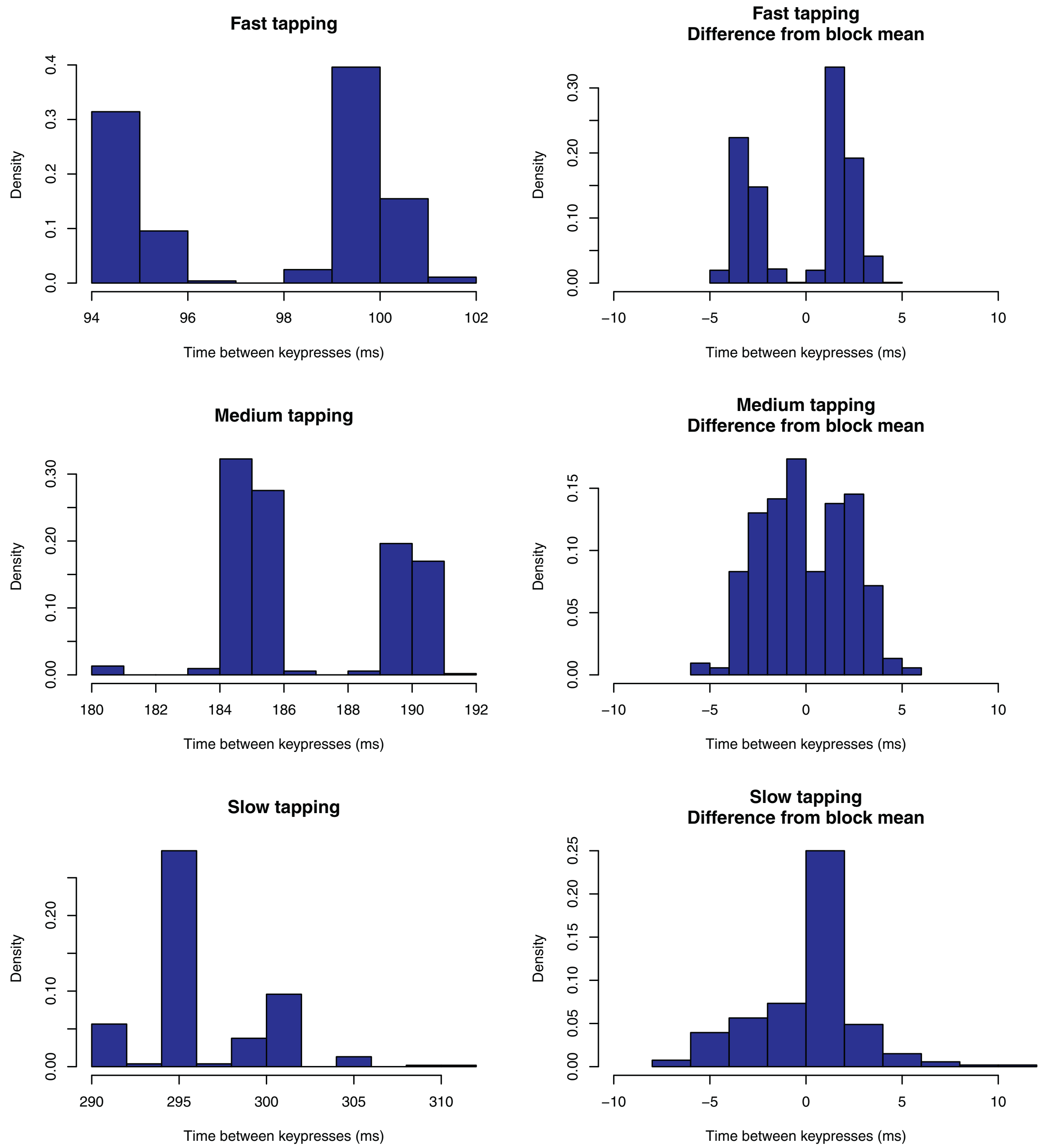

Fig. 3. Histograms of absolute and de-centered inter-tapping times for three different tapping speeds. These results show that keyboard input timing is precise to roughly 5 ms.

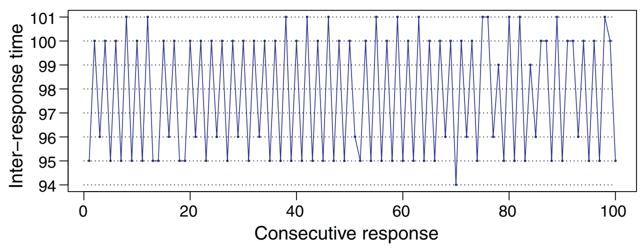

Fig. 4. The inter-response time from 100 consecutive responses. Times vacillate between 95 ±1, and 100 ± 1 ms, indicating input events are processed in 5-ms epochs.

produced similar profiles, but with a slightly longer polling time of around 8 ms.

6.3. Display timing

A third aspect of timing precision that is important for researchers is the display timing. Typical video monitors operate at a constant refresh rate that it as least 60 Hz (and some can be twice as fast), meaning that the precision with which an experimenter can control the presence of an image on the screen can be as long as in 16 ms units (although possibly as low as 8 ms). Special devices, even some dating back to Helmholtz's Laboratory, allow greater precision (see Cattell, 1885, who described a device capable of 0.1 ms precision), but modern computers with standard display devices are typically much less precise because of the limitations of consumer hardware.

On a digital computer, images get displayed when a memory buffer associated with video display is updated by a computer program. Then, the screen location corresponding to that memory location gets redrawn on the next display cycle. Under typical conditions, this update happens asynchronous to the program's timing cycle, and so there can be uncertainty (on average, half an update cycle) about when a stimulus actually appears on the screen, when it is removed, and how long it remains on-screen. In this mode, a `Draw()` command issued by PEBL will draw pending graphical changes to the memory buffer and return immediately, regardless of the timing of the screen update cycle, and the effect will appear on screen on the next video update cycle.

By default, PEBL running on Windows uses this mode, as we have found it to be most robust to different hardware and driver combinations of our users. However, when the directx video driver is used (by specifying '–driver directx' as a command-line option) and PEBL is run in full-screen mode, PEBL uses a double-buffered display. In a double-buffered display scheme, all imagery first gets written to the 'back' buffer rather than directly to the screen's buffer, and at the end of each drawing cycle the current video buffer gets swapped with the back buffer. This enables the entire screen contents to be updated on each step, preventing 'tearing' errors in which only part of the screen memory has been update when the screen is redrawn. In addition, the `Draw()` command used in PEBL will not return until the video swap has been completed. This allows one to synchronize the program to the video swap, as long as several `Draw()` commands are issued consecutively. It also allows for more consistent control over the onset, offset, and duration of stimuli, which can be displayed for a specific number of `Draw()` cycles.

To examine the precision with which one can control `Draw()` cycles, we developed another script that randomly sampled 1000 values between 1 and 16 inclusively, and then issued exactly that many `Draw()` commands (immediately preceded by three `Draw()` commands. The results of `GetTime()` were recorded immediately before and immediately after the target commands, for comparison to the programmed display times. For each programmed cycle count, we compared the observed duration of the stimulus to the smallest one observed, and computed the number of ms each stimulus was over the minimum. Results showed that 392/1000 values were at exactly the minimum value for the programmed draw cycles (roughly 16.667 times the number of programmed draw cycles), 601 were 1 ms longer, and 7 were 2 ms longer. In this test, none of the trials 'missed' a draw update and ended up drawing an extra display cycle (which would be at least 16 ms longer). In general, experimenters can use this method to have control over stimulus duration, and they can also measure the stimulus duration to determine whether any particular trials failed to draw the precise expected duration.

Several caveats must be understood about drawing precision with PEBL. First, Version 2.0 of the SDL gaming library uses hardware acceleration for most drawing operations, which has the potential to provide better synchronization to the display monitor. PEBL Version 0.13 does not use this version of the library, but future versions may, and so aspects of synchronizing to the screen may differ in the future. Second, synchronization has only been well-tested using the directx driver on windows. Other platforms may not support display synchronization in the same way. Finally, it can be difficult to assess, without special instrumentation hardware, the exact timeline of a particular draw cycle. Consequently, the exact time at which the stimulus appears on the screen may be systematically offset from the time at which the `Draw()` function returns, allowing the current time to be recorded. Thus, applications requiring exact stimulus-locked registration of response time (e.g., for synchronizing EEG or EMG signals) may be systematically offset from the actual beginning of an image on the screen.

7. Legal and ethical issues of open source testing software

A number of other legal and ethical issues are involved with developing and distributing open source psychological tests. Importantly, tests are intellectual property that are at times protected by copyright, patent, and trademark law (see Mueller, 2012a,b). As discussed in great detail by Feldman and Newman (2013), copyright has been used extensively by test publishers to prevent the free use of tests in medical settings, sometimes overstepping the actual materials that are protected by copyright.

7.1. Copyright protection of tests

Typically, only the specific implementation or expression of a test is covered by and protected under copyright law. This can include the specific language, instructions, imagery, sounds, source code, and binary code of a computerized or paper test. Although some of these may be usable under fair use doctrine, in general, imagery and like material with copyrights owned by others cannot be used and redistributed without permission (and sometimes payments) to the original rights holder. Thus, truly open tests must develop and distribute completely new stimuli, and all of the tests in the PEBL Test Battery do so.

Currently, many tests that are distributed freely by researchers in the academic community are copyrighted intellectual property owned by their university or publisher, which might someday be leveraged should it be commercially beneficial (see Newman and Feldman, 2011). Consequently, there is a difference between "freeware", which is given without charge, and Free software, which explicitly grants the rights to use, modify, and reproduce. The danger of freeware tests is that the community will invest resources in such a test, enhancing its value through replication and validation, enabling the original rights holder to capitalize on this enhanced value by charging rent. As demonstrated by Feldman and Newman (2013), this has already happened for the Mini-Mental State Exam (Folstein et al., 1975), a popular neurological screening test.

The other side of this coin is that many researchers develop tests using images, text, or sounds that are copyrighted by others (perhaps found via internet searches), and so their freeware tests may actually violate copyright law. It might be considered fair use to conduct laboratory studies using such stimuli, but researchers might possibly commit copyright infringement when those tests are shared with or licensed to others, or example stimuli are used in publications. Thus, we have been careful to avoid adding tests to the PEBL Test Battery that have unsourced imagery or test questions.

7.2. Patent protection of neurobehavioral tests

Copyright is only one kind of intellectual property law that can encumber neurobehavioral tests. The concepts, workings and mechanics of a test cannot typically be protected by copyright, but these aspects of intellectual property may be protected by patents. Patents have a much more limited timeframe, and are much more difficult to obtain, but when a test is protected by a patent, it typically means that any work-alike test that is produced, used, or distributed is subject to the patent. Many patent applications and granted patents cover tests of cognitive function. For example, U.S. Patent No. US 6884078 B2 (2005) covers a "Test of parietal lobe function and associated methods", and involves a visual test of shape and color identification. Similarly, Clark et al. (2009) noted "The Information Sampling Task is subject to international patent PCT/GB2004/003136 and is licensed to Cambridge Cognition plc." Although this patent has not been granted, the pending application has led us to avoid distributing a working version of the information sampling task.

Thankfully, most practicing researchers choose not to patent their tests, and thus give up their potential exclusivity for the greater good and the ability of others to freely test, reproduce, and bolster their findings. Researchers who choose not to patent their tests often benefit more from having their test widely used, replicated, and normed by the research community, than they would have if they had filed a patent.

7.3. Trademark protection of neuropsychological tests

In addition, trade names and trademarks are important intellectual property that grant exclusive use to those engaged in commerce with a name. Users and developers of open and free tests need to be aware of these issues in order to avoid confusion. Marks used in commerce are often unregistered, creating an area of uncertainty around the use of specific names to refer to tests. For example, it is unclear whether Berg (1948) registered a trademark for the popular "Wisconsin Card Sorting Test", which was first referred to as "University of Wisconsin Card-Sorting Test" (Grant and Berg, 1948). As early as 1951 researchers had begun referring to it as the now-common "Wisconsin Card Sorting Test" and "WCST" (Grant, 1951; Fey, 1951). At some later point, Wells Printing company began printing and distributing a paper version of the test, using the name "Wisconsin Card Sorting Test" for many years before filing for a registered mark for a paper version in 1998.[2] Heaton (1981) published a standardization manual for the test, revisions of which have later been published by Psychological Assessment Resources (PAR), Inc, who distributes both paper and computerized versions of the test. Consequently, it is unclear whether the term 'WCST' is a trademark, and this might only be decidable in a court of law. Because of this confusing state of affairs, we have typically referred to the PEBL implementation of the test as "Berg's Card-Sorting Test".

Distributing open source testing software also involves navigating ethical issues. For example, APA Ethics Code, 9.07, states "Psychologists do not promote the use of psychological assessment techniques by unqualified persons, except when such use is conducted for training purposes with appropriate supervision." (see http://www.apa.org/ethics/code/index.aspx). In contrast, the Section 2B of the General Public License (GPL), the open source license which PEBL and most of the PEBL Test Battery is licensed under specifically prohibits restraint on disclosure of software licensed under it, stating "You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License." (http://www.gnu.org/licenses/gpl-2.0.html). Although the spirit of these are at odds with one another, they are both motivated by an appeal to ethics. We must point out that the APA statement codifies an opposition to the desire of researchers to publish their results and methods, and the desire of the scientific community to examine and test methods used to derive results.

Because test descriptions are typically published in publicly accessible journals which traditionally happened to be held only in academic libraries, the guidelines essentially rely on test security through obscurity. We believe that the PEBL Test Battery does not promote the use of assessment techniques among unqualified persons any more than does every university library, and every clinical journal available therein or through on-line indexes such as google; or the many tests available as downloadable content for commercial experimentation systems; or the numerous neuropsychological testing and assessment manuals available at many libraries and bookstores, possibly including the DSM-5 (American Psychiatric Association, 2013), and certainly less than popular mental testing websites such as lumosity.com that include many psychological tests as cognitive training games. In deference to the guidelines, many commercial psychological testing companies restrict access of different tests to credentialed clinicians. However, restricting access has its own ethical dilemmas: if the community relies solely on commercial providers to enforce restricted access to credentialed practitioners, this tends to give those commercial actors monopolies, driving up health care costs and preventing access to many possible patients, caregivers, and researchers who might benefit. This especially includes organizations with limited funding or little support for psychological services, including schools, community health-care centers, prisons, and most health-care systems in less-developed countries.

Apparent violations of Ethical Code 9.07 have caused substantial uproar among the clinical psychological community, especially following an incident where the Wikipedia entry for the Rorschach test was updated to include the public domain imagery used in the test (see Cohen, 2009). Importantly, most, if not all, tests in the PEBL battery were first developed as research tasks, and later repurposed for clinical assessment. This includes widely-used continuous performance tests (initially designed to study vigilance and sustained attention, but commonly used to assess attention deficits), the Wisconsin Card Sorting Test (inspired by paradigms used to study rhesus monkeys in the behaviorist tradition, but now commonly used to measure executive function; cf. Eling et al., 2008); the Iowa Gambling Task (originally designed to test the somatic marker hypothesis); the Trail-making Test (originally testing mental flexibility, and later used as a measure sensitive to brain damage and age-related decline); the Tower of London (initially testing basic research questions of how damage to frontal cortical areas impair planning), Sperling's partial report method (original used to study iconic memory; later determined by Lu et al., 2005, to be diagnostic of cognitive decline related to Alzheimer's disease), and so on (Lezak et al., 2012). Each of these tests has substantial non-clinical usage, and so Ethics Code 9.07 appears at odds with the open scientific community which allowed these tests to be developed in the first place.

8. Summary

In this paper, we described the PEBL system and PEBL Test Battery. A detailed listing of the PEBL Test Battery was included, along with a previously unavailable review of past research methods these tests were based on. We described several tests that establish the timing precision capable with PEBL, and concluded with a discussion of the legal and ethical issues involved in open science and the open testing movement.

Acknowledgements

We would like to thank the many researchers around the world that have helped to improve PEBL by promoting, translating, using, and citing it, and through documenting and reporting bugs. B.J.P. is supported by the National Institute of Drug Abuse (L30 DA027582).

Appendix. Example choice experiment programmed using PEBL

The following experiment presents 20 O and X stimuli, instructing the participant to type they key corresponding to each one.

[2] According to trademark registration #2320931, Ser # 75588988, the mark had been in use since at least 1970. See http://tess2.uspto.gov/bin/showfield?f=doc&state=4004:wm2i54.3.1.

Mean response time is calculated and displayed at the end. The code here should be saved in a text file with a .pbl file extension, and can be run via the PEBL launcher.

```
##  Demonstration PEBL Experiment
##  Simple implementation of a choice RT task
##  Shane T. Mueller

#Every experiment needs a Start() function to initiate
define Start(p)
{
  #Define the stimuli as a list of characters
  stimuli <- ["X","O"]

  #How many times should each stimulus be presented?
  reps <- 10

  #We need to create a window object:
  window <- MakeWindow()

  #Give simple instructions in a message box:
  MessageBox("In this test, you will see either the letter 'X' or
  'O' on the screen. Type the key corresponding to the letter you
  see", window)

  #Create a stimulus sequence
  stimuli <- Shuffle(RepeatList(stimuli,reps))

  #Create a label to put stimuli on:
  target <- EasyLabel("+",gVideoWidth/2,gVideoHeight/2,window, 25)

  #Draw this and wait a second:

  rtsum <- 0
  loop(i,stimuli)
  {
    target.text <- "+"
    Draw()
    Wait(1000)
    Draw()
    time1 <- GetTime()
    target.text <- i
    Draw()
    resp <- WaitForKeyPress(i)
    time2 <- GetTime()
    rtSum <- rtSum + (time2-time1)
}

  MessageBox("Mean response time: "+rtSum/(reps*2) + "ms",window)
}
```

References

Aggarwal R, Mishra A, Crochet P, Sirimanna P, Darzi A. Effect of caffeine and taurine on simulated laparoscopy performed following sleep deprivation. British Journal of Surgery 2011;98:1666–72.

Ahonen B, Carlson A, Dunham C, Getty E, Kosmowski KJ. The effects of time of day and practice on cognitive abilities: the PEBL Pursuit Rotor, Compensatory Tracking, Match-to-sample, and TOAV tasks; 2012, PEBL Technical Report Series [On-line], #2012-02.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.

Anderson K, Deane K, Lindley D, Loucks B, Veach E. The effects of time of day and practice on cognitive abilities: the PEBL Tower of London, Trail-making, and Switcher tasks; 2012, PEBL Technical Report Series [On-line], #2012-04.

Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition 1994;50:7–15.

Berg EA. A simple objective technique for measuring flexibility in thinking. Journal of General Psychology 1948;39:15–22.

Berteau-Pavy D, Raber J, Piper B. Contributions of age, but not sex, to mental rotation performance in a community sample; 2011, PEBL Technical Report Series [On-line], #2011-02.

Bezdjian S, Baker LA, Lozano DI, Raine A. Assessing inattention and impulsivity in children during the Go/NoGo task. British Journal of Developmental Psychology 2009;27(2):365–83, http://dx.doi.org/10.1348/026151008X314919.

Carter R, Woldstad J. Repeated measurements of spatial ability with the manikin test. Human Factors: The Journal of the Human Factors and Ergonomics Society 1985;27(2):209–19.

Cattell JM. The inertia of the eye and brain. Brain 1885;8(3):295–312.

Cinaz B, Vogt C, Arnrich B, Tröster G. Implementation and evaluation of wearable reaction time tests. Pervasive and Mobile Computing 2012;8(6):813–21.

Clark L, Roiser J, Robbins T, Sahakian B. Disrupted reflection impulsivity in cannabis users but not current or former ecstasy users. Journal of Psychopharmacology 2009;23(1):14–22, http://dx.doi.org/10.1177/0269881108089587.

Clark DG, Kar J. Bias of quantifier scope interpretation is attenuated in normal aging and semantic dementia. Journal of Neurolinguistics 2011;24:411–9.

Cohen N. A Rorschach Cheat Sheet on Wikipedia? The New York Times 2009(July), Retrieved from http://www.nytimes.com/2009/07/29/technology/internet/29inkblot.html

Conners CK, Epstein JN, Angold A, Klaric J. Continuous performance test performance in a normative epidemiological sample. Journal of Abnormal Child Psychology 2003;31(5):555–62.

Cools R, Clark L, Owen AM, Robbins TW. Defining the neural mechanisms of probabilistic reversal learning using event related functional magnetic resonance imaging. Journal of Neuroscience 2002;22(11):4563–7.

Corsi PM. Human memory and the medial temporal region of the brain. Dissertation Abstracts International 1972;34:819B.

Crone EA, van der Molen MW. Developmental changes in real life decision making: performance on a gambling task previously shown to depend on the ventomedial prefrontal cortex. Developmental Neuropsychology 2004;25:250–79.

Croschere J, Dupey L, Hilliard M, Koehn H, Mayra K. The effects of time of day and practice on cognitive abilities: forward and backward Corsi block test and digit span; 2012, PEBL Technical Report Series [On-line], #2012-03.

Damos DL, Gibb GD. Development of a computer-based naval aviation selection test battery. Pensacola, FL: Naval Aerospace Medical Research Lab; 1986, DTIC Document ADA179997 http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA179997

Danckert J, Stöttinger E, Quehl N, Anderson B. Right hemisphere brain damage impairs strategy updating. Cerebral Cortex 2012;22(12):2745–60.

Dinges DI, Powell JW. Microcomputer analysis of performance on a portable, simple visual RT task sustained operations. Behavioral Research Methods, Instrumentation, and Computers 1985;17:652–5.

Ebbinghaus H. Memory: A Contribution to Experimental Psychology. New York: Teachers College, Columbia University; 1913.

Eling P, Derckx K, Maes R. On the historical and conceptual background of the Wisconsin Card Sorting Test. Brain and Cognition 2008;67(3):247.

Eriksen CW, Schultz DW. Information processing in visual search: a continuous flow conception and experimental results. Perception & Psychophysics 1979;25(4):249–63.

Fan J, McCandliss BD, Sommer T, Raz M, Posner MI. Testing the efficiency and independence of attentional networks. Journal of Cognitive Neuroscience 2002;14:340–7.

Feldman R, Newman J. Copyright at the bedside: should we stop the spread? Stanford Technology Law Review 2013;16:623–55.

Fey ET. The performance of young schizophrenics and young normals on the Wisconsin Card Sorting Test. Journal of Consulting Psychology 1951;15(4):311.

Fitts PM. The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology 1954;47:381–91.

Folstein MF, Folstein SE, McHugh PR. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research 1975;12(3):189–98.

Freeman GL. The relationship between performance level and bodily activity level. Journal of Experimental Psychology 1940;26:602–8.

Garavan H. Serial attention within working memory. Memory & Cognition 1998;26:263–76.

Gehring WJ, Bryck RL, Jonides J, Albin RL, Badre D. The mind's eye, looking inward? In search of executive control in internal attention shifting. Psychophysiology 2003;40:572–85.

Geller AS, Schleifer IK, Sederberg PB, Jacobs J, Kahana MJ. PyEPL: a cross-platform experiment-programming library. Behavior Research Methods 2007;39(4):950–8.

Grant DA. Perceptual versus analytical responses to the number concept of a Weigl-type card sorting test. Journal of Experimental Psychology 1951;41(1):23–9.

Grant DA, Berg E. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. Journal of Experimental Psychology 1948;38(4):404–11.

Greenberg LM, Kindschi CL, Corman CL. TOVA test of variables of attention: clinical guide. St. Paul, MN: TOVA Research Foundation; 1996.

Greenwald AG, McGhee DE, Schwartz JKL. Measuring individual differences in implicit cognition: the Implicit Association Test. Journal of Personality and Social Psychology 1998;74:1464–80.

Gullo MJ, Stieger AA. Anticipatory stress restores decision-making deficits in heavy drinkers by increasing sensitivity to losses. Drug & Alcohol Dependence 2011;117:204–10.

Hart S, Staveland L. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock P, Meshkati N, editors. Human mental workload; 1988. p. 139–83, Amsterdam: North Holland.

Heaton RK. A manual for the Wisconsin card sorting test. Western Psychological Services 1981.

Hirshman E, Bjork RJ. The generation effect: support for a two-factor theory. Journal of Experimental Psychology: Learning, Memory & Cognition 1986;14:484–94.

Huettel SA, McCarthy G. What is odd in the oddball task? Prefrontal cortex is activated by dynamic changes in response strategy. Neuropsychologia 2004;42(3):379–86.

Jaeggi SM, Buschkuehl M, Jonides J, Perrig WJ. Improving fluid intelligence with training on working memory. Proceedings of the National Academy of Sciences 2008;105(19):6829–33.

Jerison HH, Crannell CW, Pownall O. Acoustic noise and repeated time judgements in a visual movement projection task (WAUC-IR-b7-54). Wright-Patterson Air Force Base, OH: Wright Air Development Center; 1957.

Kalinowska-Łyszczarz A, Pawlak MA, Michalak S, Losy J. Cognitive deficit is related to immune-cell beta-NGF in multiple sclerosis patients? Journal of the Neurological Sciences 2012;321(1-2):43–8.

Kessels RPC, van Zandvoort MJE, Postma A, Kappelle LJ, de Haan EHF. The Corsi block-tapping task: standardization and normative data. Applied Neuropsychology 2000;7(4):252–8.

Kessels RPC, van den Berg E, Ruis C, Brands AMA. The backward span of the Corsi block-tapping task and its association with the WAIS-III digit span. Assessment 2008;15(4):426–34, http://dx.doi.org/10.1177/1073191108315611.

Kotovsky K, Hayes JR, Simon HA. Why are some problems hard? Evidence from Tower of Hanoi. Cognitive Psychology 1985;17(2):248–94.

Kötter R. A primer of visual stimulus presentation software. Frontiers in Neuroscience 2009;3(2):163.

Krause F, Lindemann O. Experiment: a Python library for cognitive and neuroscientific experiments. Behavior Research Methods 2013:1–13, http://dx.doi.org/10.3758/s13428-013-0390-6.

Lejuez CW, Read JP, Kahler CW, Richards JB, Ramsey SE, Stuart GL, Strong DR, Brown RA. Evaluation of a behavioral measure of risk-taking: The Balloon Analogue Risk Task (BART). Journal of Experimental Psychology: Applied 2002;8:75–84.

Lezak MD, Howison DB, Bigler ED, Tranel D. Neuropsychological assessment. 5th ed Oxford: New York; 2012.

Lipnicki DM, Gunga H, Belavy DL, Felsenberg D. Decision making after 50 days of simulated weightlessness. Brain Research 2009;1280:84–9, http://dx.doi.org/10.1016/j.brainres.2009.05.022.

Logan GD, Cowan WB, Davis KA. On the ability to inhibit simple and choice reaction time responses: a model and a method. Journal of Experimental Psychology: Human Perception and Performance 1984;10(2):276.

Lu Z-L, Neuse J, Madigan S, Dosher B. Fast decay of iconic memory in observers with mild cognitive impairments. Proceedings of the National Academy of Science 2005;102:1797–802.

Lyvers M, Tobias-Webb J. Effects of acute alcohol consumption on executive cognitive functioning in naturalistic settings. Addicitive Behavior 2010;35(11):1021–8.

Mackworth NH. The breakdown of vigilance during prolonged visual search. Quarterly Journal of Experimental Psychology 1948;1:6–21.

Mathôt S, Schreij D, Theeuwes J. OpenSesame: an open-source, graphical experiment builder for the social sciences. Behavior Research Methods 2012;44(2):314–24, http://dx.doi.org/10.3758/s13428-011-0168-7.

Meyer DE, Schvaneveldt RW. Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. Journal of Experimental Psychology 1971;90(2):227–34, http://dx.doi.org/10.1037/h0031564.

Mueller ST. A partial implementation of the BICA Cognitive Decathlon using the Psychology Experiment Building Language (PEBL). International Journal of Machine Consciousness 2010;2:273–88.

Mueller ST. The PEBL Manual, Version 0.13. Lulu Press; 2012a, ISBN 978-0557658176 http://www.lulu.com/shop/shane-t-mueller/the-pebl-manual/paperback/product-20595443.html

Mueller ST. Developing open source tests for psychology and neuroscience; 2012, December, Feature at opensource.com, http://opensource.com/life/12/12/developing-open-source-tests-psychology-and-neuroscience

Müller-Lyer FC. Optische urteilstäuschungen. Archiv für Physiologie 1889(Suppl):263–70.

Nelson B. The effects of attentional cuing and input methods on reaction time in human computer interaction; 2013, PEBL Technical Report Series [On-line] #2013-01.

Ness V, Arning L, Niesert HE, Stuettgen MC, Epplen JT, Beste C. Variations in the GRIN2B gene are associated with risky decision-making. Neuropharmacology 2011;61(5-6):950–6.

Newman JC, Feldman R. Copyright and open access at the bedside. New England Journal of Medicine 2011;365:2447–9.

Peirce JW. PsychoPy – Psychophysics software in Python. Journal of Neuroscience Methods 2007;162:8–13.

Perez WA, Masline PJ, Ramsey EG, Urban KE. Unified Tri-services cognitive performance assessment battery: review and methodology; 1987, DTIC Document ADA181697 http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA181697

Peterson L, Peterson MJ. Short-term retention of individual verbal items. Journal of Experimental Psychology 1959;58(3):193–8.

Phillips WA. On the distinction between sensory storage and short term visual memory. Perception and Psychophysics 1974;16:283–90.

Piper BJ. Age, handedness, and sex contribute to fine motor behavior in children. Journal of Neuroscience Methods 2011;195(1):88–91, http://dx.doi.org/10.1016/j.jneumeth.2010.11.018.

Piper BJ. Evaluation of the test-retest reliability of the PEBL continuous performance test in a normative sample; 2012, PEBL Technical Report Series [On-line], #2012-05.

Piper BJ, Li V, Eiwaz MA, Kobel YV, Benice TS, Chu AM, Olson R, Rice D, Gray H, Mueller ST, Raber J. Executive function on the psychology experiment building language tests. Behavior Research Methods 2012;44(1):110–23.

Piquet M, Balestra C, Sava S, Schoenen J. Supraorbital transcutaneous neurostimulation has sedative effects in healthy subjects. BMC Neurology 2011;11(1):135.

Posner MI. Orienting of attention. Quarterly Journal of Experimental Psychology 1980;32:3–25.

Premkumar M, Sable T, Dhanwal D, Dewan R. Circadian levels of serum melatonin and cortisol in relation to changes in mood, sleep and neurocognitive performance, spanning a year of residence in Antarctica. Neuroscience Journal 2013:1–10, http://dx.doi.org/10.1155/2013/254090, Article 254090.

Qiu J, Helbig R. Body posture as an indicator of workload in mental work. Human Factors 2012;54:626–63, http://dx.doi.org/10.1177/0018720812437275.

Salthouse TA, Toth J, Daniels K, Parks C, Pak R, Wolbrette M, et al. Effects of aging on the efficiency of task switching in a variant of the Trail Making Test. Neuropsychology 2000;14:102–11.

Sternberg S. High-speed scanning in human memory. Science 1966;153(3736):652–4.

Stins JF, Polderman TJC, Boomsma DI, de Geus EJC. Conditional accuracy in response interference tasks: evidence from the Eriksen flanker task and the spatial conflict task. Advances in Cognitive Psychology 2007;3(3):389–96.

Straw AD. Vision Egg: an open-source library for realtime visual stimulus generation. Frontiers in Neuroinformatics 2008;2., http://dx.doi.org/10.3389/neuro.11.004.2008, 4.

Towse J, Neil D. Analyzing human random generation behavior: a review of methods used and a computer program for describing performance. Behavior Research Methods 1998;30(4):583–91, http://dx.doi.org/10.3758/BF03209475.

Treisman A. Preattentive processing in vision. Computer Vision, Graphics, and Image Processing 1985;31(2):156–77.

Troyer AK, Leach L, Strauss E. Aging and response inhibition: normative data for the Victoria Stroop Test. Aging, Neuropsychology & Cognition 2006;13(1):20–35, http://dx.doi.org/10.1080/138255890968187.

Tumkaya S, Karadag F, Mueller ST, Ugurlu TT, Oguzhanoglu NK, Ozdel O, Atesci FC, Bayraktutan M. Situation awareness in Obsessive-Compulsive Disorder. Psychiatry Research 2013;209(3):578–88, http://dx.doi.org/10.1016/j.psychres.2013.02.009.

Ulrich NJ III. Cognitive performance as a function of patterns of sleep; 2012, PEBL Technical Report Series [On-line], #2012-06.

Waugh NC, Norman DA. Primary memory. Psychological Review 1965;72(2):89–104.

Wilkinson RT, Houghton D. Portable four-choice reaction time test with magnetic tape memory. Behavior Research Methods 1975;7(5):441–6.

Worthy DA, Hawthorne MJ, Otto AR. Heterogeneity of strategy usage in the Iowa Gambling Task: a comparison of win-stay lose-shift and reinforcement learning models. Psychonomic Bulletin & Review 2013;20:364–71.

Yamaguchi M, Proctor RW. Multidimensional vector model of stimulus-response compatibility. Psychological Review 2012;119:272–303.