

ARTÍCULO DE INVESTIGACIÓN

<http://dx.doi.org/10.14482/psdc.38.3.153>



Validación de Experimentos Cognitivos con PEBL y Wundt's Lab

*Validation of Cognitive Experiments
with PEBL and Wundt's Lab*

JUAN FRANCISCO MUÑOZ OLANO

<https://orcid.org/0000-0001-9055-2586>

MICHELLE NATALIA TORRES VILLATE

<https://orcid.org/0000-0003-4209-0342>

YEFERSON ANTONIO ZAVALA FONSECA

<https://orcid.org/0000-0002-5311-5456>

Universidad Pedagógica y Tecnológica de Colombia

Correspondencia: juan.munoz04@uptc.edu.co



Resumen

Actualmente las herramientas tecnológicas están reemplazando los métodos tradicionales en diversos campos de la psicología, un fenómeno de alto impacto; por esta razón se realizó un análisis de las características de validez y confiabilidad de pruebas experimentales computarizadas empleando un estudio psicométrico instrumental que contó con 267 participantes que realizaron 589 aplicaciones pareadas de pruebas, en tanto que la mayoría llevó a cabo más de dos tareas cognitivas. Mediante un análisis discriminante, se identificaron cuatro pruebas computarizadas PEBL y tres pruebas computarizadas Wundt's Lab, consistentes con medidas cognitivas validadas, a diferencia de tres pruebas PEBL y cuatro pruebas Wundt's Lab, que no mostraron tal evidencia. Adicionalmente, se encontraron convergencias entre los desempeños de pruebas Mental Rotation de PEBL y Wundt's Lab, así como asociaciones pareadas entre desempeños de pruebas de memoria operativa verbal y visoespacial, de atención y de percepción del movimiento. Finalmente, se evidenciaron asociaciones relevantes entre los resultados de los participantes en las pruebas y la variable "sexo". Los resultados apoyan la existencia de características psicométricas en siete pruebas computarizadas, haciendo relevante su uso en diversos campos de la evaluación cognitiva.

Palabras claves: PEBL, Wundt's Lab, validez, experimentos, memoria, atención, percepción.

Abstract

Currently, technological tools are replacing traditional methods in many fields of psychology, a phenomenon of high impact. For this reason, validity and reliability characteristics analysis of computerized experimental tests were carried out, using an instrumental psychometric study, which had 267 participants taking part in 589 paired applications of tests. Through a discriminant analysis, four computerized PEBL tests and three computerized Wundt's Lab tests were identified, consistent with validated cognitive measures, as opposed to three PEBL tests and four Wundt's Lab tests, which did not show such evidence. Additionally, convergences were found between PEBL and Wundt's Lab's Mental Rotation test performances, as well as paired associations between verbal and visuospatial working memory, and attention

Citación/referenciación: Muñoz Olano, J., Torres Villate, M. & Zavala Fonseca, Y. (2021). Validación de Experimentos Cognitivos con PEBL y Wundt's Lab. *Psicología desde el Caribe*, 38(3), 368-392.

and movement perception test performances. Finally, relevant associations were shown between the results of the participants in the tests and the sex variable. Results support the existence of psychometric characteristics in seven computerized tests, making their use relevant in various fields of cognitive evaluation.

Keywords: PEBL, Wundt's Lab, validity, test, experiment, memory, attention, perception.

Introducción

Una amplia oferta de programas computarizados ha surgido como alternativa a las pruebas de lápiz y papel (Germine et. al, 2012; López, Sanmartín y Méndez, 2013). Sin embargo, se ha planteado que estas alternativas aún no cumplen con todos los criterios de validez y confiabilidad necesarios para ser usados indiscriminadamente en ámbitos aplicados de la psicología (Buelow y Suhr, 2009). No obstante, ya son varias las investigaciones al respecto que han medido procesos cognitivos como las funciones ejecutivas, memoria de trabajo y atención, utilizando recursos tecnológicos computarizados (Ruano et. al, 2016; de Paula, Malloy-Diniz y Romano-Silva, 2016; Dougherty et. al, 2010; Ruiz, Guevara y Hernández, 2008). Incluso se han incluido estrategias computarizadas en la enseñanza de procesos cognitivos (Muñoz-Olano y Acevedo-Triana, 2018).

Diversos *softwares* han sido empleados para obtener medidas de funciones ejecutivas (*v.g.*, atención, abstracción y flexibilidad mental), así como de lenguaje, procesamiento espacial, función sensoriomotora, y memorias verbal y espacial (Reitan y Woltson, 1985); incluso en las labores de rehabilitación cognitiva (Suárez de Oliveira, Trezza, Busse & Filho, 2014) y tratamiento a población con demencia (Dougherty et. al, 2010).

El *software* de acceso libre *Psychology Experimental Building Language v 2.0* (PEBL-Wiki, 2014) es una plataforma gratuita construida en lenguaje de programación C++ abierto y modificable según la necesidad del investigador, ya que utiliza una licencia abierta (GNU Public License 2.0). Es de resaltar la existencia de pocos estudios que acrediten la validez de PEBL, aunque sí existen varios estudios que prueban la fiabilidad de sus medidas (Ramos, 2017; de Paula, Malloy-Diniz

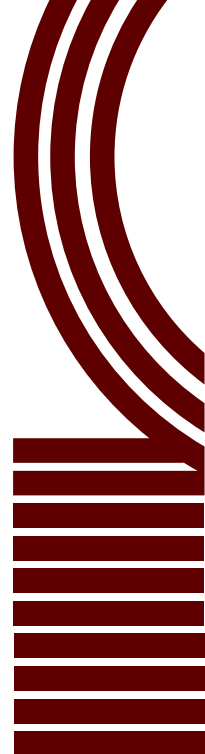
y Romano-Silva, 2016 Muller, 2015; Ramos, Guevara, Bolaños y Ramos, 2015; Piper & Mueller, 2014).

Pruebas PEBL como *Digit Span* han evidenciado correlaciones con otras pruebas de memoria de hasta $r = .68$, y buenos niveles de correlación test- retest (Dikmen, Heaton, Grant & Temkin, 1999). Como han planteado Dikmen et. al (1999), correlaciones Spearman y Pearson menores o iguales a .7 se han llegado a interpretar como aceptables; correlaciones entre .3 y .7 se han interpretado con niveles de aceptación intermedios; y valores de correlación por debajo de .3 se han interpretado como inaceptables. Incluso pruebas como *Time Wall Task*, también de PEBL, han demostrado una mayor precisión con la práctica (Piper, Mueller, Talebzadeh & Ki, 2016).

El *software De La Rosa Research Wundt's Lab* (Wundt's Lab) es un programa licenciado multimedia, creado por una empresa líder en mercado psicológico en Colombia; cuenta con 43 experimentos cognitivos, que incluye versiones de algunas reconocidas pruebas en procesos cognitivos específicos (Research, 2017). Este *software* no cuenta con ningún estudio que analice sus características de fiabilidad y validez en sus medidas.

El estado del arte plantea ventajas en la utilización de pruebas computarizadas; se puede encontrar que requieren un menor tiempo de aplicación, han probado fiabilidad en la medición de tiempos de respuesta, permiten una mayor rapidez en la calificación de las pruebas, y finalmente, presentan un potencial de adaptación a las nuevas generaciones (Suárez de Oliveira et. al, 2014; Milne, Culverwell, Guss, Tuppen & Whelton, 2008). Además, las pruebas computarizadas han mostrado una considerable capacidad de procesar una gran cantidad de variables y datos, proveen puntajes y retroalimentación en tiempo real y reducen el costo de la aplicación de pruebas, considerando las necesidades de poblaciones especiales (Abal et. al, 2008).

Así mismo, otra ventaja que presentan las pruebas computarizadas frente a las pruebas psicológicas de lápiz y papel es que no se encuentran regidas por extensos tiempos de aplicación y cansancio ante las mismas; por este motivo, estos instrumentos se han beneficiado especialmente de la evolución de las pruebas informáticas (Gur et. al, 2010) y las facilidades que estas presentan. Adicionalmente, la extensión de pruebas computarizadas puede aportar a la solución del problema de replicabilidad de experimentos psicológicos



que se presenta con mayor frecuencia respecto a otras ciencias y disciplinas (Acevedo-Triana, López-López y Cárdenas, 2014; Cumming, 2014; Loannidis, 2005), al permitir la obtención de más datos que los validen. Finalmente, estas herramientas ofrecen importantes alternativas de acceso remoto y virtualidad a través de Internet (Ruano et. al, 2016), sobre todo en estos tiempos de cuarentena por los efectos mundiales de la pandemia por COVID-19.

No obstante, los estudios en el campo también han resaltado las limitantes de las pruebas cognitivas computarizadas. Primero, la escasa posibilidad de acceso a datos cualitativos, que solo se pueden obtener por medio de la interacción directa con cada participante (Brandwayn, Macías y Restrepo, 2015; Tsiakas, Gagnon, Awad & Messier, 2004); segundo, los efectos del aprendizaje al repetir el uso de las pruebas, algo que puede afectar la sensibilidad de estos experimentos cognitivos para medir ciertos aspectos relacionados con adquirir habilidades cognitivas (Abal et. al, 2008); y finalmente, las dificultades inherentes al uso de nuevas tecnologías, como la falta de familiaridad con las mismas (Ruano et. al, 2016; Suárez de Oliveira et. al, 2014).

Por estas razones, el objetivo de este estudio fue evidenciar criterios de validez en las medidas obtenidas de las aplicaciones de pruebas PEBL y Wundt's Lab. Para lograr este objetivo, este estudio inició con una validez de contenido (Yaghmale, 2003) orientada a identificar el fenómeno cognitivo evaluado por cada prueba. De este modo, fue acordado entre jueces analizar los desempeños en experimentos cognitivos de PEBL y Wundt's Lab que en teoría miden los mismos fenómenos psicológicos, tales como la toma de decisiones, la atención, la memoria operativa, la imaginación mental y la percepción del movimiento.

Así, se evidenciaron características de validez y confiabilidad de catorce pruebas computarizadas, siete PEBL y siete Wundt's Lab. Entre las pruebas PEBL estuvieron *Iowa Gambling Task*, *Brown Peterson*, *Change Detection Task*, *Mental Rotation*, *Digit Span*, *Corsi Blocks Task* y *Time Wall Task*; entre las pruebas Wundt's Lab estuvieron, *Risky Decisions*, *Brown Peterson*, *Change Detection*, *Mental Rotation*, *Memory Span*, *Position Error*, y finalmente, *Apparent Motion*. Todas las pruebas fueron presentadas con sus respectivas versiones en español.

Método

Consistió en una investigación multivariada de tipo correlacional (Coolican, 1997), con un diseño psicométrico instrumental (Montero & León, 2007),

empleado para obtener parámetros de validez discriminante (Hogan, 2015), y pruebas de asociación entre variables, entre los datos obtenidos de aplicaciones de las pruebas PEBL y Wundt's Lab. Esta metodología ha sido ampliamente utilizada en el campo de la validación de cuestionarios psicológicos (Pineda, 2016), y fue aplicada para evaluar el rendimiento de estas pruebas cognitivas computarizadas.

Participantes

El estudio midió el desempeño en los experimentos cognitivos por parte de la población de jóvenes en la ciudad de Tunja (Boyacá). Si bien esta población es la más común en estudios universitarios, todavía existe una importante carencia de datos sobre aplicaciones de las pruebas PEBL (Piper & Mueller, 2014) y Wundt's Lab, incluso con esta población.

El tamaño muestral fue de $n = 267$, con base en una población estimada en 48 609 jóvenes (municipio de Tunja, 2015). El cálculo muestral se hizo con un poder de .95, un valor de error de probabilidad del .01, y un posible tamaño del efecto desde $ES = .3$, considerando una significancia de las asociaciones entre variables con nivel menor o igual a .05, a dos colas. Estos cálculos se hicieron con el *software* estadístico *G*Power* (2018). La selección de la muestra se realizó por conveniencia, con participantes que formaron parte del estudio, por la accesibilidad y proximidad con los investigadores (Coolican, 1997). Como criterios de inclusión se establecieron, primero, que sus edades estuvieran entre 18 y 29 años de edad; y segundo, a través de entrevista estructurada se desestimó que los participantes pudieran presentar limitaciones físicas (p. ej., discapacidad visual o auditiva, falta de sueño o efectos de medicamentos) o cognitivas (p. ej., diagnóstico de déficit cognitivo, déficit de atención o deterioro de alguna actividad mental). Finalmente, la muestra estuvo compuesta por participantes entre 18 años y 26 años ($M = 20,73$; $R = 8$), todos estudiantes universitarios, con una considerablemente mayor participación de estudiantes de psicología ($n = 155$), y la menor participación de otras carreras, como Enfermería ($n = 24$), Medicina ($n = 14$), y algunos pocos estudiantes de otras carreras, como Licenciatura en Matemáticas, Psicopedagogía, Derecho, Ingeniería Electrónica, Licenciatura en Educación Física, Medicina Veterinaria, Ingeniería Civil y Química. Si bien los participantes viven en la ciudad de Tunja, en una alta proporción son procedentes de otros municipios del departamento

de Boyacá (Colombia), tales como Sogamoso, Duitama, Paipa, Samacá, Santa Rosa, Ventaquemada, Villa de Leyva, Chiquinquirá, entre otros.

Instrumentos

Mediante un análisis de validez de contenido inicial realizado con los criterios de Argibay (2006) fue acordado entre jueces expertos clasificar los experimentos cognitivos. Evaluando protocolos e instrucciones de las pruebas, estas fueron pareadas de las siguientes formas.

La prueba Brown-Peterson consiste en un experimento original de los años cincuenta que mide la capacidad de la memoria de trabajo. Creado por John Brown, Lloyd y Margaret Peterson (Peterson y Peterson, 1959). Consiste en recordar trigramas no morféminos, es decir, sin significado unitario. Por ejemplo, una secuencia como “B, P, M”, mientras se calcula hacia atrás la resta de tres unidades numéricas, desde un número inicial, por ejemplo “396”. Es decir, 396 – 393 – 390 – 387... en adelante. En la versión PEBL, el experimento cuenta con 48 ensayos, de constante dificultad. En la versión Wundt´s Lab apenas cuenta con 20 ensayos.

La prueba Change Detection (Detección al Cambio) consiste en el uso de capacidades de atención endógena para encontrar un cambio en una escena visual (PEBL Change Detection Task -Wiki, 2013). La prueba genera el conocido fenómeno de la ceguera al cambio (Simons & Levin, 1998). En PEBL, la prueba Change Detection cuenta con veinte ensayos, cada uno con condiciones cambiantes en estímulos visuales, por su forma, tamaño y localización, estando además en competencia por la atención una gran variedad de estimulaciones, donde el objetivo es localizar cada estímulo cambiante en cada ensayo, entre escenarios con milisegundos de parpadeo atencional (Shapiro, Arnell & Raymond, 1997). En Wundt´s Lab, la prueba Change Detection cuenta con veintitrés ensayos, con escenarios de fotografías realistas, separadas en algunas ocasiones por parpadeo atencional, escenas ante las cuales el participante debe indicar si hubo o no cambio en algún elemento de la imagen.

La prueba Mental Rotation (Rotación mental) tiene su origen en los estudios de Shepard y Metzler (Shepard, 1978), consistentes en identificar la igualdad entre dos imágenes especulares, estando una de ellas posiblemente rotada en su eje en comparación con la otra. Estos estudios dieron origen a la investigación en simulación mental por imágenes (Smith y Kosslyn, 2008). La prueba Mental

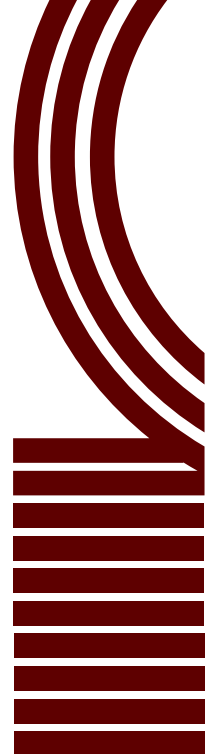
Rotation de PEBL cuenta con la amplia cantidad de 132 ensayos de comparación entre posibles imágenes especulares con rotación; estas imágenes se presentan de manera bidimensional. Por el contrario, la prueba Mental Rotation de Wundt 's Lab cuenta con imágenes de objetos posiblemente especulares y en rotación, que son tridimensionales. Esta prueba incluye veinte ensayos, en los cuales el participante determina si hay una imagen especular alterada por el efecto de la rotación.

La prueba Digit Span de PEBL se encarga de medir memoria verbal a corto plazo (PEBL - Digit Span Wiki, 2013). La prueba se fundamenta en el descubrimiento del almacenamiento acústico-verbal de las memorias recientes, razón por la cual requiere un ensayo vocal del recuerdo que supera el tiempo de almacenamiento sensorial, segundos después de la aparición de una serie de dígitos (Sperling, 1960; Baddeley, 1999a). La prueba puede llegar a quince ensayos, dependiendo del rendimiento del participante, dado que un bajo rendimiento puede llevar solo hasta nueve u once ensayos. En Wundt 's Lab, esta prueba se conoce como Memory Span, con un número de 12 ensayos.

La prueba Bloques de Corsi (o por su nombre en inglés, Corsi Blocks Task) es una prueba psicológica que evalúa la memoria de trabajo a corto plazo viso-espacial (Kessels, Zandvoort, Postma, Kappelle & de Haan, 2000). Fue creada por Philip Michael Corsi (1972) con el fin de estudiar el papel del lóbulo temporal medial en el recuerdo humano. La prueba consiste en recordar la secuencia de iluminación de cubos dispuestos en diferentes coordenadas, que incrementan en número y en patrones de secuencia de iluminación seriada.

La versión PEBL puede llegar a tener 17 ensayos, con una dificultad en incremento, aunque desempeños inferiores llegan a los 15 o 13 ensayos (PEBL Corsi Blocks Test Wiki, 2019). Una prueba que también evalúa la memoria reciente viso-espacial es la prueba Position Error (Posición del error) de Wundt 's Lab, consistente en veinte ensayos, en los que el participante debe memorizar secuencias de dígitos como, por ejemplo, "FKLMRTX", para luego indicar con precisión el lugar de cada dígito en la secuencia.

La prueba Iowa Gambling Task (Prueba de apuestas de Iowa) consiste en una simulación de toma de decisiones monetarias, introducida por Bechara, Damasio, Tranel y Anderson (1998), de la Universidad de Iowa. Si bien la prueba tuvo como función inicial identificar problemas en la toma de decisiones de pacientes



con lesiones orbitofrontales que comprometen su inhibición de respuesta (Bechara, 2004), hallazgos han demostrado la utilidad de la misma para medir impulsividad en las respuestas de participantes sin lesiones cerebrales (Weller, Levin & Bechara, 2010).

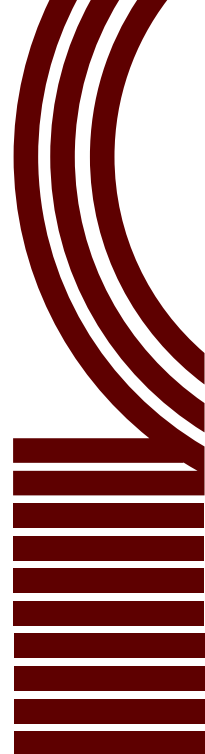
La prueba de PEBL muestra cuatro mazos de cartas, indicando que algunas elecciones significarán ganar dinero, mientras otras perder dinero. Unos mazos llevan a elecciones óptimas con frecuencia, pero a pérdidas infrecuentes. El participante debe distribuir de manera óptima sus elecciones, sin caer en la impulsividad en las mismas; la prueba cuenta con alrededor de cien ensayos, en los que se mide la optimización de la elección y la posible impulsividad de la misma. La prueba Risky Decisions de Wundt's Lab consta de doce ensayos, en los que el participante debe elegir entre varias opciones frente a un problema de riesgo de pérdida de dinero hipotético. Esta prueba no cuenta con estudios previos sobre su validez y su confiabilidad.

La prueba Time Wall de PEBL (Muro de Tiempo) consiste en una evaluación de la forma como se percibe el movimiento aparente. Es una tarea básica de estimación de un tiempo requerido, en el cual un objeto en movimiento desaparece detrás de un muro, y el participante debe juzgar el tiempo exacto en que el objeto habría alcanzado un destino final (PEBL Time Wall Wiki, 2013). Así, en la prueba debe estimarse su velocidad de movimiento en caída, hasta terminar en un espacio, al final del muro, durante 20 ensayos.

Apparent Motion de Wundt's Lab (Movimiento aparente de Wundt's Lab) consiste en una prueba basada en el fenómeno *Phi* de persistencia retinada, descubierto originalmente en 1912 por Max Wertheimer (Shiffman, 2004). Mediante un método gradual de medición de umbral de percepción del movimiento aparente, el participante durante 23 ensayos, ajusta velocidades de rotación de círculos concéntricos, hasta lograr la ilusión óptica de percepción de un movimiento constante a una velocidad *Phi* de alrededor de 200 ms.

Procedimiento

Consistió en un proceso de ocho meses, durante el cual asistieron los participantes voluntarios, convocados por anuncios e invitaciones, a las aplicaciones llevadas a cabo, principalmente en el Laboratorio de Procesos Psicológicos (LAPROPSIC) de la Escuela de Psicología de la Universidad Pedagógica y Tecnológica de Colombia (UPTC). El laboratorio cuenta con alrededor de cinco computadores



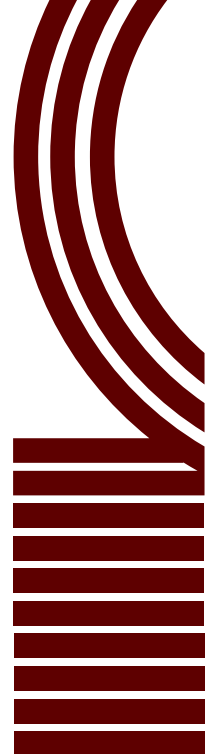
personales con licencias requeridas. Además, se empleó un equipo portátil para la realización de algunas aplicaciones por fuera del laboratorio, en condiciones también óptimas de aislamiento.

Cada participante eligió participar llevando a cabo dos, cuatro o hasta seis pruebas pareadas, siendo así como se analizaron 589 resultados totales obtenidos con todas las pruebas. Cada aplicación por participante tuvo un control de contrabalanceo, recomendado para experimentos y el uso de pruebas psicológicas (Martin, 2008), consistente en aleatorizar el orden en que los participantes llevaban a cabo varias de las pruebas PEBL y Wundt's Lab que elegían realizar. Es de resaltar siempre se informaba sobre el tiempo promedio requerido para realizar dos (*e.j.*, entre 10 y 15 minutos), cuatro (*e.j.*, entre 20 y 30 minutos) o hasta seis pruebas (*e.j.*, entre 40 y 50 minutos), esta última opción tomada por pocos participantes que evidenciaban no presentar agotamiento en las tareas. Una práctica común de las aplicaciones era cronometrar los tiempos en cada experimento para tener una idea promedio de los minutos requeridos.

Así, combinando un ambiente propicio y condiciones de selección de participantes y condiciones de aplicación, se logró el control de variables extrañas y de confusión, como dificultades cognitivas, efectos de agotamiento y sueño, falta de interés en la tarea, o falta de claridad en instrucciones y aplicaciones, fuentes recomendadas a tener presente en estos estudios (Coolican, 1997).

Como parte del procedimiento, constantemente se llevaron a cabo análisis preliminares de los datos obtenidos, así como controles de las condiciones experimentales de aplicación requeridas en lo ético.

Con respecto al análisis de los datos obtenidos, como primer paso se llevaron a cabo pruebas de normalidad de las distribuciones obtenidas, al medir con cada prueba cada proceso cognitivo (*e.j.*, toma de decisiones, imaginación mental, memoria de trabajo); como segundo paso, se obtuvo un análisis de fiabilidad por dos mitades, recomendado para datos obtenidos transversalmente (Prieto y Delgado, 2010); como tercer paso, se hizo un análisis loglineal de asociación multivariada, con el fin de evidenciar las interacciones entre los desempeños en las pruebas y variables ilustrativas como el tipo de *software* de la prueba, el sexo y la edad de los participantes; como cuarto paso, se realizaron análisis bivariados relevantes para comprobar relaciones dadas en el desempeño de las pruebas; y finalmente, se llevó a cabo un análisis de clasificación discriminante, con el fin



de identificar el comportamiento de las pruebas para medir atributos y procesos cognitivos.

Consideraciones éticas

Las condiciones de aplicación fueron evaluadas por el comité de la Dirección de Investigaciones de la Universidad Pedagógica y Tecnológica de Colombia, y se determinó un riesgo mínimo para los participantes (Resolución 8430 de 1993). Las aplicaciones también contaron con la elaboración de un consentimiento informado, que propende por el cumplimiento de las normas éticas de medición y experimentación en psicología (Ley 23 de 1982; Ley habeas data de 2008; Consideraciones éticas para la investigación científica, 2014).

Resultados

Tres pruebas de Wundt's Lab y cuatro pruebas de PEBL evidenciaron una distribución normalizada; además, con condiciones de igualdad de varianzas entre las comparaciones medidas entre las pruebas de Mental Rotation (MR) de Wundt's Lab y MR de PEBL, así como entre las pruebas de Risky Decisions (RD) de Wundt's Lab y Iowa Gambling Task (IGT) de PEBL. Los resultados reposan en plataforma de acceso libre (Muñoz, 2020).

Un análisis de fiabilidad por dos mitades fue requerido, dada la obtención de medidas en un solo evento (Prieto y Delgado, 2010). Tomando en cuenta el total de aplicaciones ($n = 589$), un α de *Guttman* = .80 corroboró la confiabilidad de las medidas de desempeño obtenidas con las pruebas de ambos *softwares*.

Pruebas Wundt's Lab y PEBL

Las pruebas MR, RD y Apparent Motion (AM) de Wundt's Lab evidenciaron distribuciones normalizadas. Por el contrario, las distribuciones de pruebas Wundt's Lab Brown Peterson - BP ($n = 45$), Change Detection- CD ($n = 42$), Memory Span- MS ($n = 42$), y Position Error - PE ($n = 45$), no presentaron distribución normalizada; fueron las pruebas BP, MR, IGT y DS de PEBL, las que evidenciaron pruebas normalizadas desde este *software*. Por el contrario, las distribuciones de pruebas PEBL, CD (40), Corsi Blocks Task - CBT (45) y Time Wall - TW (35) presentaron distribuciones alejadas de igualdades de varianza y normalidad (véase tabla 1).

■ **Tabla 1.** Distribuciones normalizadas encontradas en pruebas PEBL y Wundt´s Lab

	N	M	Mdn	Dt	Min	Max	IC- 95%		SW		
							Li	l.s	e	gl	p
Porcentajes de aciertos Brown Peterson PEBL (sin outliers)	44	63,77	64,58	13,22	33,33	87,50	59,75	67,79	,971	44	,337
Porcentajes de aciertos Rotación Mental Wundt´s Lab	44	60,34	60	11,43	40	85	56,86	63,81	,952	44	,064
Porcentajes de aciertos Rotación Mental PEBL (sin outliers)	41	56,07	53,91	10,76	36,72	78,91	52,68	59,47	,946	41	,052
Total de Riesgo Risky Decisions Wundt´s Lab	43	42,24	41,67	15,14	8,33	75	37,58	46,91	,959	43	,124
Pérdida Total IGT de PEBL (sin outliers)	38	66,06	66,53	8,50	49,26	85,68	63,26	68,85	,985	38	,872
Porcentajes de aciertos Digit Span de PEBL	45	39,73	41,18	12,89	17,65	64,71	35,83	43,63	,961	45	,135
Porcentaje de aciertos en Apparent Motion de Wundt´s Lab (sin outliers)	39	54,12	55	13,54	24,38	89,38	49,72	58,51	,968	39	,317

Nota: elaboración propia. N (número), M (media aritmética), Mdn (Mediana), Min (valor mínimo), Max (valor máximo), dt (desviación típica), IC (intervalo de confianza, con Li - límite inferior, y Ls, límite superior), SW (prueba Shapiro Wilk, con e - probabilidad de error, gl - grados de libertad, y p - nivel de significancia).

Asociaciones entre variables

Un análisis loglineal de Poisson evidenció asociaciones parciales entre variables como la edad y el sexo de los participantes, con sus niveles de desempeño y la utilización del respectivo *software* Wundt´s Lab o PEBL, al realizar las pruebas cognitivas.

Para el caso de las pruebas BP, el análisis loglineal tuvo bondad de ajuste ($LR = .72$, $gl = 6$, $p = .99$). Los modelos 1 y 2 resultaron significativos ($\chi^2 = 65.4$, $gl = 15$, $p < .05$; $\chi^2 = 21.7$, $gl = 11$, $p = .03$), evidenciando asociaciones parciales entre el nivel de acierto en las pruebas con el tipo de *software* empleado, así como con el sexo

y la edad de los participantes ($x^2 = 6.4$, $gl = 1$, $p = .011$). Así, las puntuaciones de acierto tendieron a ser más altas en la prueba BP de Wundt's Lab comparadas con las puntuaciones en la prueba BP de PEBL. Estas diferencias impactaron en los resultados relacionados con la edad y el sexo, dada una pequeña ventaja de los hombres entre 21 y 26 años para obtener un mayor desempeño que los mismos hombres entre 18 y 20 años. Adicionalmente, las mujeres entre 18 y 20 años presentaron ventaja en el nivel de acierto comparadas con aquellas entre 21 y 26 años de edad.

Con respecto a las pruebas MR, el análisis loglineal también presentó bondad de ajuste ($LR = 8.3$, $gl = 10$, $p = .603$). El modelo 1 resultó significativo ($x^2 = 26.9$, $gl = 15$, $p = .027$), evidenciando asociaciones parciales en la interacción entre el nivel de acierto en la prueba y el sexo ($x^2 = 4.58$, $gl = 1$, $p = .032$), sin que se presentara una asociación entre el nivel de acierto de los participantes y el tipo de *software* empleado. De hecho, mediante Análisis Lineal Generalizante se comprobó que los niveles de acierto en MR de Wundt's Lab y PEBL presentaron igualdad de varianzas ($F = 3.1$, $gl = 1$, $p = .81$). No obstante, en promedio, en el nivel de acierto se presentó una pequeña ventaja para los hombres que diferencia sus puntuaciones de las de las mujeres ($t = 3.2$, $gl = 86$, $p = .002$).

Las pruebas RD de Wundt's Lab e IGT de PEBL, PE de Wundt's Lab y CBT de PEBL, CD de Wundt's Lab y PEBL, y MS de Wundt's Lab y DS de PEBL evidenciaron asociaciones pareadas que demuestran la diferencia obtenida en los niveles de aciertos de los participantes, dependiendo del *software* de la prueba. Así, los resultados pareados de RD de Wundt's Lab e IGT de PEBL muestran que las puntuaciones de IGT en aciertos suelen ser considerablemente más altas que las de RD ($x^2 = 52.1$, $gl = 1$, $p < .05$); la comparación pareada de PE de Wundt's Lab con CBT de PEBL evidencia una diferencia similar, aunque menor, dado que las puntuaciones de acierto en PE tienden a ser más altas que las de CBT ($x^2 = 5.5$, $gl = 1$, $p = .02$); adicionalmente, se evidenció una diferencia en el nivel de acierto, según si la prueba es de CD de Wundt's Lab o de PEBL ($x^2 = 12.5$, $gl = 1$, $p < .05$); finalmente, el análisis pareado entre MS de Wundt's Lab y DS de PEBL mostró que las puntuaciones en aciertos en MS fueron considerablemente más altas que las obtenidas con DS ($x^2 = 67.9$, $gl = 1$, $p < .05$).

En la figura 1 pueden observarse las diferencias entre los promedios y valores marginales dados en los niveles de acierto obtenidos por los participantes en las

diferentes pruebas, evidenciando aquellas con mayores y menores diferencias en sus puntuaciones.

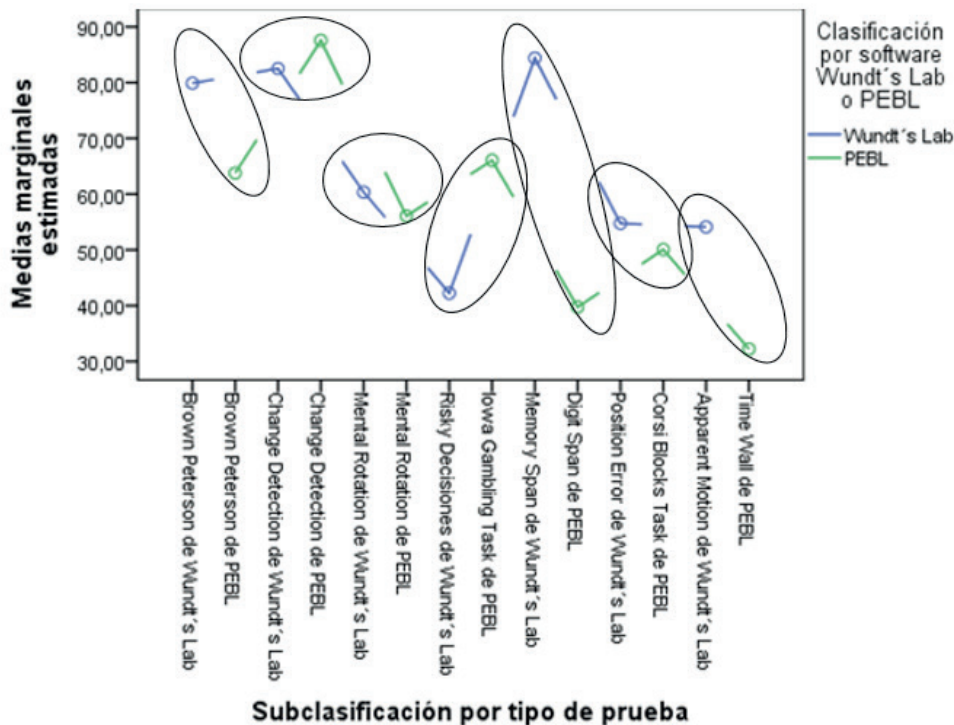


Figura 1. Diferencias de medias entre niveles de acierto de los participantes en cada prueba

Comparaciones bivariadas

Se presentaron alrededor de cinco comparaciones bivariadas significativas, principalmente por análisis no paramétricos, dada la participación de distribuciones no normalizadas en estas comparaciones.

Los análisis de los resultados obtenidos con las pruebas BP de Wundt's Lab y PEBL evidencian ser muestras relacionadas, además porque las puntuaciones de BP Wundt's Lab tienden a ser considerablemente más altas que en la prueba BP de PEBL ($W^+ = 4.6$, $DE = 82.8$, $p < .05$). También las pruebas CD de Wundt's Lab y PEBL evidenciaron rangos con signo de muestras relacionadas ($W^+ = 2.49$, $DE = 77.1$, $p < .05$). Los resultados de aplicaciones de las pruebas MS de Wundt's Lab y DS de PEBL también evidenciaron relaciones mutuas ($W^+ = 5.65$, $DE = 79.9$, $p < .05$). Como último análisis no paramétrico, se evidenció una relación entre los resultados de pruebas de AM de Wundt's Lab y TW de PEBL ($W^+ = 4.99$, $DE = 63.65$,

$p < .05$). Como comparación paramétrica, se evidenció una independencia significativa entre los resultados obtenidos en pruebas RD de Wundt 's Lab e IGT de PEBL, siendo $t = -9.134$, $gl = .37$, $p = < .05$, IC [-28.46 - 18.12].

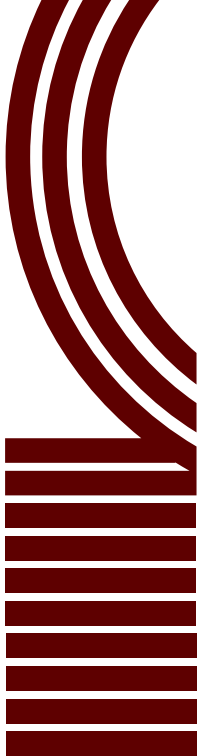
Agrupación de los niveles de acierto en pruebas

Se realizó un análisis factorial, no obstante, las distribuciones no cumplieron con los requisitos para el mismo. Para analizar condiciones de validez de constructo se empleó el análisis de clasificación discriminante. Se comprobó una agrupación del desempeño en las pruebas por niveles de esfuerzo cognitivo. Esta clasificación es coherente con puntos de vista teóricos sobre el esfuerzo mental requerido en tareas con demandas cognitivas (Kahneman, 2012). Así, del total de aplicaciones $n = 589$ se encontraron 170 casos ponderados de aplicaciones de un menor nivel de esfuerzo cognitivo, 294 casos ponderados de un nivel óptimo de esfuerzo cognitivo, y 125 casos ponderados de aplicaciones de un nivel alto de esfuerzo cognitivo.

Inicialmente se identificó el ajuste al análisis discriminante ($M \text{ Box} = 238.8$, $F = 19.7$, $p < .05$). Presentándose además una probabilidad de .33 de pertenencia previa para cada uno de los tres grupos planteados para la agrupación.

Fueron comprobadas diferencias significativas entre las varianzas de las distribuciones de las variables clasificadoras: Medidas en pruebas Wundt 's Lab y PEBL ($\Lambda = .37$, $F = 484.4$, $gl = 2$, $p < .05$); clasificación por *software* Wundt 's Lab o PEBL ($\Lambda = .89$, $F = 34.7$, $gl = 2$, $p < .05$); y finalmente, la subclasificación por tipo de prueba ($\Lambda = .70$, $F = 123.8$, $gl = 2$, $p < .05$).

Las funciones discriminantes explicaron la clasificación de los grupos, con una $r \text{ canónica} = .816$, que explica el 99.4 % de la varianza total. De hecho, un contraste entre funciones corroboró diferencias entre los tres grupos clasificados ($x^2 = 647.6$, $gl = 6$, $p < .05$). No obstante, una primera función de clasificación diferenció más claramente a los grupos. Las funciones en los centroides clasificadas por nivel de esfuerzo cognitivo fueron: Para la función 1, menor esfuerzo cognitivo (1.99), esfuerzo óptimo (-.335) y alto esfuerzo (-1.918); mientras que para la función 2 de clasificación, menor esfuerzo cognitivo (-.074), esfuerzo óptimo (.106) y alto esfuerzo (-.148). Como puede verse en la tabla 2, se presentó una clasificación más evidente desde la función 1.



El análisis de puntuaciones promedio y marginales en las pruebas, en complemento con un análisis discriminante, confirmó la diferenciación de las pruebas por nivel de esfuerzo y desempeño requeridos. Así, un grupo considerable de pruebas permitieron un nivel de desempeño y esfuerzo cognitivo óptimos, al presentar su nivel promedio de desempeño entre el 50 % y el 70 %, niveles intermedios. Entre estas pruebas estuvieron Brown-Peterson de PEBL, Mental Rotation de Wundt's Lab y PEBL, Iowa Gambling Task de PEBL, Position Error de Wundt's Lab, Corsi Blocks Task de PEBL y Apparent Motion de Wundt's Lab; otro grupo de pruebas presentó en promedio un muy alto desempeño, llevando a considerar que su ejecución no requería un grado significativo de esfuerzo mental. Entre estas pruebas estuvieron Brown-Peterson Wundt's Lab, Change detection de Wundt's Lab y Change Detection de PEBL, y Memory Span de Wundt's Lab; y finalmente, tres pruebas presentaron en promedio niveles de desempeño bajos, dada su dificultad en ejecutarse una sola vez. Estas pruebas fueron Risky Decisions de Wundt's Lab, Digit Span de PEBL y Time Wall de PEBL.

Por último, los coeficientes lineales de las funciones discriminativas evidenciaron un aporte de todas las variables independientes (v.g., medidas en pruebas Wundt's Lab y PEBL, clasificación por *software* y subclasificación por tipo de prueba), en la clasificación de los grupos. No obstante, el mayor aporte fue dado por la variable de clasificación por *software*.

■ **Tabla 2.** Coeficientes de Fisher de la función de clasificación

	Nivel de esfuerzo cognitivo		
	Mayor esfuerzo	Esfuerzo óptimo	Esfuerzo alto
Medidas en pruebas de Wundt's Lab y PEBL	.563	.410	.292
Clasificación por software	5.180	6.930	7.537
Sub clasificación por tipo de prueba	.833	1.07	1.176

Fuente: elaboración propia.

Discusión

El objetivo de este estudio fue evidenciar los criterios de validez psicométrica de los softwares de evaluación psicológica De La Rosa Research Wundt's Lab y Psychology Experimental Building Language v 2.0, así como las posibles relaciones y asociaciones entre resultados en pruebas cognitivas; en ese sentido,

los resultados apoyan la existencia de parámetros de validez de las medidas obtenidas de algunas pruebas; con menor apoyo, pero con algunos casos significativos, se encontraron relaciones y asociaciones entre los resultados de pruebas computarizadas con objetivos psicológicos similares.

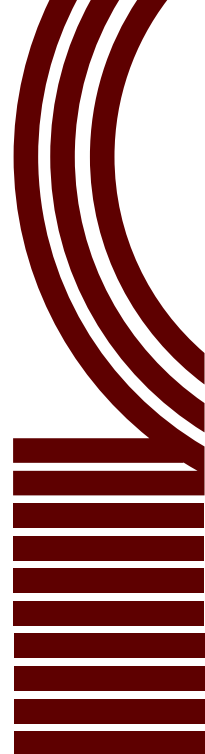
El estudio arrojó un índice alto de confiabilidad. Esto en concordancia con lo encontrado por Piper et. al (2011 y 2015) y Buelow y Suhr (2009). En cuanto a los criterios de validez, por análisis discriminante, se encontró que siete pruebas están en el nivel óptimo de esfuerzo cognitivo, lo cual evidencia que este grupo de tareas miden correctamente el constructo psicológico en el que están fundamentadas.

No se logró determinar correlaciones convergentes entre las parejas de pruebas. Sin embargo, se encontraron asociaciones parciales en seis parejas, en las cuales se relacionaron las variables sexo, edad, rendimiento y tipo de *software*; además de encontrar rangos de desempeño con una relación proporcional entre cada una de las parejas, en la mayoría de las pruebas, exceptuando la pareja de Iowa Gambling Task de PEBL (IGT) y Risky Decisions de Wundt's Lab (RD).

La prueba Brown-Peterson (BP) de Wundt's Lab presentó un nivel de acierto mayor respecto a su contraparte de PEBL. Esto pudo ser ocasionado por el número de ensayos de cada prueba, e incluso, en mayor medida porque la prueba de Wundt's Lab presenta interferencias de carácter verbal, mas no aritméticas y auditivas, como se plantea en el experimento original y en estudios posteriores (Tsiakas et. al, 2003).

En las pruebas de Mental Rotation (MR) se evidenció una similitud en la forma como tanto la prueba de PEBL como la de Wundt's Lab lograron funcionar para generar simulación del movimiento e imagería mental. De hecho, ambas pruebas entregaron una igualdad de varianzas en los desempeños de los participantes. A pesar de diferencias entre las pruebas, como en el número de ensayos, en la aleatorización de sus ensayos, solo presente en PEBL, y en la diagramación tridimensional (Wundt's Lab) o bidimensional (PEBL) de sus imágenes especulares, ambas pruebas probaron medir los mismos fenómenos de simulación mental.

No obstante, en las pruebas de MR se encontró una pequeña ventaja en rendimiento para los hombres, de hecho, probada por análisis loglineal. Aunque no es el objeto de este estudio plantear diferencias categóricas por sexo,

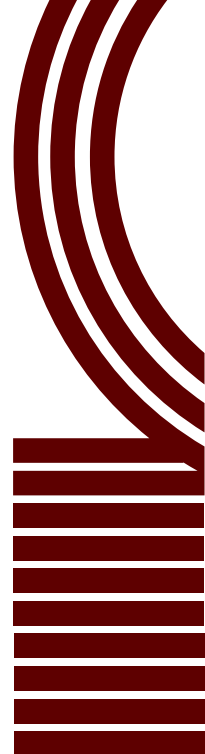


resaltamos hechos soportados por análisis estadísticamente significativo, y que también, en el caso de las diferencias por sexo en rotación mental, son hechos encontrados por otros investigadores, como el caso de Blum, Auné y Attorresi (2015; ver también Gil-Verona et. al, 2003), cuando su estudio concluyó que los hombres tienen una mayor capacidad para girar mentalmente objetos, en el reconocimiento de formas y en la representación bidimensional de objetos tridimensionales. Cabe además mencionar que diferentes investigaciones apoyan empíricamente la teoría de una capacidad espacial influida por factores ambientales como jugar a videojuegos (Pietsch y Jansen, 2012), una variable que pudo beneficiar a algunos hombres sobre algunas mujeres, que no estuvo controlada por el estudio.

En cuanto a las pruebas de Change Detection (CD), se encontró una relación proporcional en los rangos de puntuaciones PEBL y Wundt's Lab, aunque con un mayor nivel de acierto en CD de Wundt's Lab respecto a PEBL; de hecho, en CD de PEBL se evidenció que factores de atención exógena, como el color, posición, tamaño, novedad y complejidad de estímulos, fueron componentes que afectaron la medida precisa del funcionamiento atencional, dado que la prueba falla en no tener un tiempo de respuesta limitado para el esfuerzo atencional. Algo que explicaría por qué la prueba se ubicó con aquellas que solo requieren un bajo esfuerzo cognitivo para su alto desempeño.

En lo referente a las pruebas que evaluaban la toma de decisiones, no se encontró ningún tipo de asociación entre las mismas, posiblemente debido a que Risky Decisions (RD) de Wundt's Lab está fundamentada en la teoría de rechazo al riesgo (Dunn, Lutes & Risko, 2016), mientras que la prueba Iowa Gambling Task (IGT) de PEBL se basa en la teoría del *Risk Seeking* (Koban, 2008). Es necesario resaltar que la prueba IGT se ubicó en un nivel de desempeño óptimo, algo acorde con los estudios de Gansler, Jerram, Vannorsdall y Schretlen (2011), Buelow y Suhr (2009) y Piper et. al (2016) sobre IGT, mientras que RD se ubicó en un desempeño de alto esfuerzo cognitivo. En las aplicaciones de RD, algunas dificultades propias de la validez interna del experimento se hicieron presentes, relacionadas con la comprensibilidad de sus preguntas y tareas de decisión.

Con respecto a pruebas que miden memoria viso-espacial, se comprobó que tanto Corsi Blocks Task (CBT) de PEBL y Position Error (PE) de Wundt's Lab se ubicaron en un nivel óptimo de esfuerzo mental requerido; sin embargo, los participantes contaron con un puntaje superior en PE respecto a CBT. No

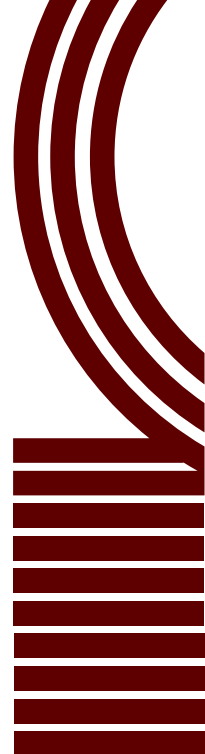


obstante, ambas pruebas evidenciaron medir con precisión el ejecutivo central y su complemento de almacén viso-espacial, planteado en el modelo de Baddeley y Hitch (Smith y Kosslyn, 2008). De hecho, en estas dos tareas fue evidente el efecto de primacía, concepto originalmente propuesto por Herman Ebbinghaus (Gluck, Mercado y Myers, 2009), al notar la constante interferencia del recuerdo de primeras secuencias sobre secuencias posteriores.

En las pruebas Time Wall (TW) de PEBL y Apparent Motion (AM) de Wundt's Lab se demostró que TW es una prueba dirigida principalmente a medir la percepción de velocidad de un objeto en movimiento aparente (Smith y Kosslyn, 2008), y esto la convierte en una prueba de alto esfuerzo cognitivo; mientras que AM se centra en el efecto ilusorio de la velocidad *Phi*, un movimiento aparente más intuitivo y sencillo de identificar (Shiffman, 2004). De hecho, las medidas de ambas pruebas no presentaron ninguna relación, dado además que PE clasificó como una prueba de esfuerzo óptimo, y TW de alto esfuerzo cognitivo. Incluso, TW suele requerir prácticas repetidas para su adecuado dominio (Piper et. al, 2011), algo comprobado en este estudio que lleva a considerar varias aplicaciones para que TW pruebe su consistencia interna.

Memory Span (MS) de Wundt's Lab presentó puntuaciones más altas que Digit Span (DS) de PEBL, debido a la presencia constante de recordatorios visuales sobre los dígitos a evocar verbalmente en la prueba. Por el contrario, DS exige una evocación absoluta de la memoria auditiva, sin señales o ayudas adicionales, un proceso fundamental para el logro de memorias auditivas verbales recientes, pasado el medio segundo de registro inicial de la estimulación (Baddeley, 1999b). De hecho, durante la aplicación se logró evidenciar que las personas utilizaban recursos nemotécnicos (v.g., vocalizar) para recordar los estímulos presentados visual y auditivamente.

La validez interna de pruebas experimentales suele ser sensible a problemas relacionados con cómo funcionan los experimentos (Martin, 2008). Pruebas de memorias recientes verbales y viso-espaciales requieren de la cantidad suficiente de ensayos para lograr un nivel óptimo de esfuerzo mental. Precisamente, la diferencia entre pruebas como BP de PEBL, con suficientes ensayos, y BP de Wundt's Lab, sin suficientes. Además, cada proceso cognitivo requiere de sus componentes más fundamentales. Por ejemplo, BP de PEBL tiene un funcionamiento óptimo porque genera interferencias entre los primeros y los segundos recuerdos en las pruebas, al igual que la prueba CBT.



Para algunos procesos, el número de ensayos parece no requerir ser amplio, como sucedió en la medición de la imaginación mental por rotación de imágenes. Las pruebas de MR en Wundt's Lab y PEBL difieren considerablemente en su número de ensayos, y midieron de manera similar las capacidades de rotación mental de objetos geométricos. No obstante, pruebas como BP, CBT e IGT de PEBL evidenciaron que conllevan a una mejora progresiva en su desempeño cuando los participantes logran en ellas un alto nivel de ejecución en múltiples ensayos, contrario a quienes no lo logran, dado que evidencian abruptas desmejoras en las fases intermedias y finales de estas pruebas.

Las variables de confusión, consistentes en creer que se mide un proceso cognitivo cuando la prueba o experimento causa otros efectos (Martin, 2008), también son fuentes de invalidación. Por ejemplo, en el caso de la prueba CD de Wundt's Lab, la presencia de imágenes sin parpadeo atencional afectó considerablemente la medición de este fenómeno, totalmente indispensable para generar los fallos en el espacio atencional requerido (Gusev, Mikhaylova & Utochkin, 2014). Igualmente, las condiciones de esfuerzo atencional deben permitir tiempos límite de decisión, algo que afecta las medidas óptimas de CD de PEBL, al no tener tiempos mínimos de ejecución, alterando las condiciones requeridas para el esfuerzo demandado en detectar disyunciones en relaciones entre grupos de estímulos (Bateman, Ngiam & Birney, 2018).

Conclusiones y recomendaciones

Tal vez solo el tiempo y la constante investigación en el campo logren concluir qué tanto los experimentos computarizados logran solucionar problemas de validez en medidas psicológicas y cognitivas en diferentes poblaciones, y ante una variedad de requerimientos evaluativos. Por el momento, este estudio apoya la evidencia que plantea que aunque existen una variedad de pruebas computarizadas, y varias de ellas no parecen pasar las pruebas psicométricas, otras sí lo logran. Además, brinda posibilidades que se escapan a las pruebas de lápiz y papel, en cuanto a facilidades de aplicación y acceso remoto, estandarización de instrucciones y pasos, precisión de medidas de respuestas y procesos, y potenciales aplicaciones a través de nuevas tecnologías.

Se recomienda que nuevos estudios se enfoquen particularmente en las pruebas que miden de manera óptima el proceso cognitivo. La validación de las mismas no termina con este estudio, y requiere ampliarse a otras poblaciones y campos

de aplicación. Además, las pruebas identificadas como alto esfuerzo cognitivo podrían seguirse probando, y determinar si sus medidas pueden mejorar con la repetición y el dominio. Cabe resaltar ciertos procesos cognitivos requieren de habilidades y aprendizajes constantes.

Referencias

- Abal, F., Blum, G. D., Aguerri, M. E., Galibert, M. S., Atorresi, H. F. & Lozzia, G. S. (2008). Test informatizados: Nuevos desafíos prácticos y éticos para la evaluación psicológica. *SUMMA Psicológica UST*, 6(1), 135-148. doi: 10.18774/448x.2009.6.57
- Acevedo-Triana, C. A., López-López, W. y Cárdenas, L. F. (2014). Recomendaciones en el diseño, la ejecución y la publicación de investigaciones en Psicología y ciencias del comportamiento. *Revista Costarricense de Psicología*, 33(2), 155-177. Recuperado de: <http://www.rcps-cr.org/openjournal/index.php/RCPs/article/view/48>
- Argibay, J. C. (2006). Técnicas psicométricas. Cuestiones de validez y confiabilidad. *Subjetividad y Procesos Cognitivos*, 8, 15 – 33. Recuperado de: <http://www.redalyc.org/articulo.oa?id=339630247002>
- Baddeley, A. (1999a). Percepción y recuerdo. En A. Baddeley, *Memoria Humana: Teoría y Práctica*. Madrid: McGraw-Hill.
- Baddeley, A. (1999b). Las imágenes visuales y la agenda viso-espacial. En *Memoria Humana: Teoría y Práctica*. Madrid: McGraw-Hill.
- Bateman, J. E., Ngiam, W. X. Q. & Birney, D. P. (2018). Relational encoding of objects in working memory: Changes detection performance is better for violations in group relations. *Plos One, Open Acces*. Recuperado de: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0209207>
- Bechara, A. (2004). The Role of Emotion in Decision-Making: Evidence from Neurological Patients with Orbitofrontal Damage. *Brain and Cognition*, 30-40. doi: 10.1016/j.bandc.2003.04.001
- Bechara, A., Damasio, H., Tranel, D. & Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *J. Neurosci.*, 18(1), 428-437. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6793407/>
- Blum, G. D., Auné, S. & Atorresi, H. F. (2015). Desempeño de varones y mujeres en relación con el razonamiento visuoespacial. Influencia de la velocidad de respuesta, la educación y la presencia de rotación en los ítems. *Revista Mexicana de Psicología*, 32(1), 16-26.
- Brandwayn, N., Macías, R. y Restrepo, D. E. (2015). Comparación en el desempeño de pruebas neuropsicológicas sistematizadas y de lápiz y papel en una muestra de estudiantes de psicología de la Pontificia Universidad Javeriana sede Bogotá. (Tesis de grado en Psicología). Recuperado de: <https://repository.javeriana.edu.co/bitstream/handle/10554/21363/BrandwaynBricenoNatalia2015.pdf?sequence=1&isAllowed=y>

- Buelow, M. y Suhr, J. (2009). Validez de construcción de la tarea de juego de Iowa. *Neuropsychology Review*, 19(1), 102-114. doi: 10.1007/s11065-009-9083-4
- Coolican, H. (1997) *Métodos de Investigación y Estadística en Psicología* (2ª ed.). México: Manual Moderno.
- Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* (Ph.D.). McGill University.
- Cumming, G. (2014). The new statistics: why and how. *Psychological Science*, 25(1), 7-29. doi:10.1177/0956797613504966
- De Paula, J. J., Malloy-Diniz, L. F. & Romano-Silva, M. A. (2016). Reliability of working memory assessment in neurocognitive disorders: A study of the Digit Span and Corsi Block-Tapping tasks. *Revista Brasileira de Psiquiatria*, 38(3), 262-263. Doi: <https://doi.org/10.1590/1516-4446-2015-1879>
- Dikmen S, S., Heaton, R. K., Grant, I. & Temkin, N. R. (1999). Test-retest reliability and practice effects of the expanded Halstead-Reitan neuropsychological test battery. *Journal of the International Neuropsychological Society*, 5, 346 - 336. <http://dx.doi.org/10.1017/S1355617799544056>
- Dougherty, J. H., Cannon, R. L., Christopher, C. R., Hall, L., Hare, F., Carr, E., Dougherty, A., Janowitz, J. & Arunthamakun, J. (2010). The computerized self test (CST): An interactive, Internet accesible cognitive screening test for dementia. *Journal of Alzheimer's Disease*, 20, 185-195. doi: 10.3233/JAD-2010-1354
- Dunn, T., Lutes, D. & Risko, E. (2016). Metacognitive evaluation in the avoidance of demand. *Journal of Experimental Psychology: Human perception and performance*. *Open Science Framework*. Recuperado de: <https://osf.io/mfcxd/>
- Gansler, Jerram, Vannorsdall & Schretlen (2011). Comparing Alternative Metrics to Assess Performance on the Iowa Gambling Task. *Journal of Clinical and Experimental Neuropsychology* 33(9), 1040-1048. Doi: 10.1080/13803395.2011.596820
- Germine, L., Nakayama, K., Duchaine, B., Chabris, C., Chaterjje, G. & Wilmer, J. (2012). *Is the Web as good as the laboratory? Comparable performance of the web and the laboratory in cognitive / perceptive experiments*. Cambridge: Psychonomic Bulletin & Review. doi: 10.3758/s13423-012-0296-9
- Gil-Verona, J. A., Macías, J. A., Pastor, J. F., de Paz, F., Barbosa, M., Maniega, M. A., Román, J. M., López, A., Álvarez-Alfageme, I., Rami-González, L. y Boget, T. (2002). Diferencias sexuales en el sistema nervioso humano. Una revisión desde el punto de vista psiconeurobiológico. *International Journal of Clinical Health Psychology*, 3(2), 351-361.
- Gluck, M. A., Mercado, E. y Myers, C. E. (2009). *Aprendizaje y Memoria: del cerebro al comportamiento*. Bogotá: McGraw-Hill.
- Gur, R., Richard, J., Hughett, P., Calkins, M., Macy, L. & Bilker, W. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods* 187(2), 254-262. doi: 10.1016/j.jneumeth.2009.11.017
- G*Power. (2018) *Statistical Power Analyses for Windows*. Universitat Dusseldorf. Recuperado de: <http://www.gpower.hhu.de/>

- Gusev, A. N., Mikhaylova, O. A. & Utochkin, I. S. (2014). Stimulus determinants of the phenomenon of change blindness. *Psychology in Russia*, 7(1), 122-134. doi: 10.11621/pir.2014.0112
- Hogan, T. P. (2015). *Prueba psicológicas: Una introducción práctica* (2ª ed.). Mexico: Manual Moderno.
- Kahneman, D. (2012). *Pensar rápido, pensar despacio*. Bogotá: Pinguin Random.
- Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J. & de Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and Normative Data. *Applied Neuropsychology*, 7(4), 252-258. doi: 10.1207/s15324826an0704_8. PMID 11296689
- Koban, C. D. (2008). sesgos cognitivos en la toma de decisiones. *International Journal* 1(1), 68-73. <https://doi.org/10.21500/20112084.968>
- Ley de habeas data de 2008. Recuperado de: <http://wp.presidencia.gov.co/sitios/normativa/leyes/Documents/Juridica/Ley%201266%20de%2031%20de%20diciembre%202008.pdf>
- Ley número 23 de 1982 sobre derechos de autor. Recuperado de: <http://derechodeautor.gov.co/documents/10181/182597/23.pdf/a97b8750-8451-4529-ab87-bb8216odd226>
- Loannidis, J. P. A. (2005) Why Most Published Research Findings Are False. *PLoS Med*, 2(8), e124. doi: <https://doi.org/10.1371/journal.pmed.0020124>
- López, R., Sanmartín, P. y Méndez, F. (2013). Revisión de las evaluaciones adaptativas computarizadas (CAT). *Educación y Humanismo* 16(26), 27-40.
- Martin, W. (2008). *Psicología Experimental: Cómo hacer experimentos en Psicología*. Cenage Leraning.
- Milne, A., Culverwell, A., Guss, R., Tuppen, J. & Whelton, R. (2008). Screening for dementia in primary care: a review of the use, efficacy and quality of measures. *International Psychogeriatrics* 20(5), 1-16. doi: 0.1017/S1041610208007394
- Montero, I. & León, O. G. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847-862. Recuperado de: <https://www.redalyc.org/pdf/337/33770318.pdf>
- Municipio de Tunja (2015). *Política Pública de Juventud*. Disponible en: <http://186.116.13.48/obsocial/documentos/POLITICA%20PUBLICA.pdf>
- Muñoz-Olano, J. F. y Acevedo-Triana, C. A. (2018). *Texto Guía en el Laboratorio de Procesos Psicológicos*. Editorial Universidad Pedagógica y Tecnológica de Colombia.
- Muñoz, J. (2020). Validation of cognitive computerized experiments [archivo de datos de acceso libre]. *Open Science Framework*. doi 10.17605/OSF.IO/RP3D7. OSFHome. <https://osf.io/rp3d7/>
- PEBL Change Detection Task -Wiki (2013). Change Detection Task. Disponible en: http://pebl.sourceforge.net/wiki/index.php/Change_Detection_task
- PEBL Corsi Blocks Test, Wiki (2019). Corsi Block Tapping Test. Disponible en: https://en.wikipedia.org/wiki/Corsi_block-tapping_test

- PEBL – Digit Span Wiki (2013). Digit Span. Retrieved from: http://pebl.sourceforge.net/wiki/index.php/Digit_Span
- PEBL Test Battery. (2014, Junio 24). *PEBL Test Battery*. The PEBL Test Battery distributes a set of psychological tests for researchers and clinicians. The current version is 0.6. It is distributed automatically with PEBL, and on windows appears in My Documents\pebl-exp\battery. . Obtenido de Wiki: http://pebl.sourceforge.net/wiki/index.php/PEBL_Test_Battery
- PEBL Time Wall Wiki (2013). Timewall. Retrieved from: <http://pebl.sourceforge.net/wiki/index.php/Timewall>
- PEBL-Wiki (2014). PEBL Test Battery. Retrieved from: http://pebl.sourceforge.net/wiki/index.php/PEBL_Test_Battery.
- Peterson, L. R. y Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58 (3), 193–198. CiteSeerX 10.1.1.227.1807. doi:10.1037/h0049234. PMID 14432252).
- Pietsch, S. & Jansen, P. (2012). Different mental rotation performance in students of music, sport and education. *Learning and Individual Differences* 22(1):159–163. DOI: 10.1016/j.lindif.2011.11.012
- Pineda, C. A. (2016). Propiedades psicométricas de una prueba de homonegatividad. *Psicología desde el Caribe*, 33 (1), 47-65. Disponible en <http://rcientificas.uninorte.edu.co/index.php/psicologia/article/view/6545/83>
- Piper, B. y Mueller, S. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods* 30(222), 250-259. doi:<http://doi.org/10.1016/j.jneumeth.2013.10.024>
- Piper, B. , Li, V. , Elwaz, M. , Kobel, Y. , Benice, T. , Chu, A. & Mueller, S. (2011). Executive Function on the Psychology Experiment Building Language Tests. *Behavior Research Methods*, 110-123. doi: 10.3758/S13428-011-0096-6.
- Piper, B. , Mueller, S. , Geerken, A. , Dixon, K. , Kroliczak, G. O. & Miller, J. (2015). Reliability and validity of neurobehavioral function on the Psychology Experimental Building Language test battery in young adults. *PeerJ*. doi: 10.7717/peerj.1460. eCollection 2015
- Piper, B. J. , Mueller, S. T. , Talebzadeh, S. & Ki, M. J. (2016). Evaluation of the Validity of the Psychology Experiment Building Language Tests of Vigilance, Auditory Memory, and Decision Making. *PeerJ*, e1772. <http://doi.org/10.7717/peerj.1772>.
- Prieto, G. y Delgado, A. R. (2010). Fiabilidad y Validez. *Papeles del Psicólogo*, 31(1), 67-74. Disponible en: <http://www.papelesdel psicologo.es/pdf/1797.pdf>
- Ramos, C. (2017). Adaptación del Test Stroop Victoria en Estudiantes Ecuatorianos. *Revista Iberoamericana de Diagnóstico y Evaluación*, 44(2), 57-64. <https://doi.org/10.21865/RIDEP44.2.05>
- Ramos, C. A. , Guevara, Bolaños, M. E. y Ramos, D. A. (2015). Adaptación y estudio descriptivo del experimento G/No-Go en una muestra de estudiantes ecuatorianos. *Revista tecnologica ESPOL*, 28(2), 119-133. Disponible en: <http://www.rte.espol.edu.ec/index.php/tecnologica/article/download/364/243>.

- Reitan, R. M. & Woltson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation*. Tucson, AZ: Neuropsychology Press.
- Research, D. L. (2017). *Wundt's Lab*. Recuperado de: <http://www.delarosaresearch.com/detail.php?productid=139#features-tab>
- Resolución 8430 de 1993, Recuperado de: <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DIJ/RESOLUCION-8430-DE-1993.PDF>
- Ruano, L., Sousa, A., Severo, M., Alves, I., Colunas, M., Moreira, S. & Cruz, V. (2016). Development of a self-administered web-based test for longitudinal cognitive assessment. *Scientific Reports* 6:19114,1-10. Disponible en <https://www.nature.com/articles/srep19114>
- Ruiz, M., Guevara, M. A. y Hernández, M. (2008). Desarrollo computacional de pruebas para evaluar funciones ejecutivas: Hanoi y Wisconsin. *Avances en la investigación científica en el CUCBA*. Disponible en: [http://www.cucba.udg.mx/sites/default/files/publicaciones1/avances/avances2008/Biologia/CienciasAmbientales\(pp231-390\)/RuizDiazMarina\(pp383-390\)/383-390.pdf](http://www.cucba.udg.mx/sites/default/files/publicaciones1/avances/avances2008/Biologia/CienciasAmbientales(pp231-390)/RuizDiazMarina(pp383-390)/383-390.pdf)
- Simons, D. J. y Levin, D. T. (1998). Failure to detect changes to ogle during a real-world interaction. *Psychonomic Bulletin*, 5(4), 644-649. Disponible en: <https://link.springer.com/content/pdf/10.3758%2FBF03208840.pdf>
- Shapiro, K. L., Arnell, K. M. y Raymond, J. E. (1997). The attentional blink. *Trends in Cognitive Sciences*, 1(8). 291-295. doi: 10.1016/S1364-6613(97)01094-2.
- Shepard, R. N. (1978). The mental image. *American Psychologist*, 33, 125-137. <https://doi.org/10.1037/0003-066X.33.2.125>
- Shiffman, H. R. (2004). *Sensación y percepción: un enfoque integrador*. México: Manual Moderno.
- Smith, E. E. y Kosslyn, S. M. (2008). *Procesos cognitivos: modelos y bases neurales*. Madrid: Pearson Prentice Hall.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1-29. Disponible en: <https://doi.org/10.1037/h0093759>
- Suárez de Oliveira, R., Trezza, B., Busse, A. & Filho, W. (2014). Computerized test and cognitive impact of interventions. *Dement Neuropsychol* 8(2),107-111. doi: 10.1016/j.eatbeh.2016.08.012
- Tsiakas, M., Gagnon, M., Awad, N. & Messier, C. (2003). Addressing the Differences in Speed of Processing of the Intervening Calculation Task on the Modified Brown-Peterson Task. *Applied Neuropsychology*, 11(2), 103-106. doi: 10.1207/s15324826an1102_6
- Weller, J. A., Levin, I. P. & Bechara, A. (2010). Do individual differences in Iowa Gambling Task performance predict adaptive decision making for risky gains and losses? *Journal of Clinical and Experimental Neuropsychology*, 32 (2), 141-150. <https://doi.org/10.1080/13803390902881926>
- Yaghmale, F. (2003). Content validity and its estimation. *Journal of Medical Education*, 3(1), 25-27. Retrieved from: file:///C:/Users/Lenovo/Downloads/Content_validity_and_its_estimation.pdf

