



“Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0”

Cristián Brotfeld¹ y Marcela Rodríguez²

¹Centro de Desarrollo de Tecnologías de Inclusión de la Pontificia Universidad Católica de Chile, CEDETi-UC

²Universidad de La Frontera

Resumen

De acuerdo a las directrices internacionales, cada vez que se genera una nueva versión de un test que implique alguna modificación en su formato de administración, se debe examinar si esta nueva versión es equivalente (en términos de confiabilidad y validez) a la original. El objetivo del presente estudio fue generar evidencia sobre la comparabilidad de las puntuaciones obtenidas en la Batería de Propensión a la Accidentabilidad Hal2.0, administrada a 20 participantes, en su versión Tablet y en Computador (versión web). Se realizaron comparaciones de puntajes promedios y a fin de estimar el efecto aprendizaje y el efecto del formato se utilizó la técnica de regresión múltiple. Los resultados muestran que no existen diferencias estadísticamente significativas entre las puntuaciones obtenidas para las pruebas de Atención al Detalle, Atención Focalizada, Inteligencia Fluida y Personalidad. Por otra parte, se detectó una diferencia en la cantidad de aciertos y errores en la prueba de Atención Dividida, que no impacta en la puntuación total de la batería. Estos hallazgos evidencian la equivalencia entre los formatos de este instrumento, validando su utilización en ambos dispositivos.

Palabras clave: Propensión a la accidentabilidad, Test informatizados, Equivalencia.





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

El desarrollo de test informatizados ha implicado grandes cambios y avances en el área de la evaluación psicológica y la psicometría en los últimos años, permitiendo la aplicación on-line de instrumentos, el diseño y la construcción de nuevos formatos de ítems o la generación de informes automatizados, especialmente en el área de la Psicología Organizacional (Bartram, 2006; Elosua, 2021, 2022; Menéndez et al., 2008; Lozzia et al., 2009; Olea et al., 2010). Los test informatizados se caracterizan por ser contestados y puntuados en un computador y según Olea et al. (1999), para que puedan ser denominados como tales, deben evidenciar que cuentan con adecuadas propiedades psicométricas diferenciándose así de otros instrumentos que se ofrecen en internet.

Informatizar un test tiene grandes ventajas tales como un mayor control de condiciones estandarizadas (control de tiempo, instrucciones comunes, entre otras.), mayor rapidez en la obtención de resultados cuantitativos, verbales o gráficos, la posibilidad de generar informes automáticos agilizando la devolución de resultados y actualizar constantemente los datos normativos. Por otro lado, los test informatizados han permitido ampliar la semejanza entre una tarea evaluativa y los criterios que se busca predecir, por ejemplo, en la selección de personas, cubriendo un rango más amplio de atributos a evaluar (Elosua, 2022; Lozzia et al., 2009; Noyes & Garland, 2008; Olea et al., 2010).

Todas estas ventajas han impulsado la informatización de test que previamente contaban con un formato tradicional (de lápiz y papel), tarea compleja y que requiere demostrar que independientemente de la modalidad, sus propiedades psicométricas siguen siendo las apropiadas (Dadey et al., 2018; Hernández, 2022). Esto es, que mantienen los estándares de validez y confiabilidad del instrumento (Brandwayn et al., 2020), dado que la sola informatización de un test no garantiza su ajuste técnico y psicométrico (Muñiz & Hambleton, 1999).

Algunos estudios que han explorado la equivalencia de un test entre su formato tradicional de lápiz y papel y uno informatizado, han generado evidencia variada. Por ejemplo, algunos han encontrado diferencias en las puntuaciones atribuibles a la forma de presentación de las instrucciones, componentes gráficos del test o en la acción de contestar las actividades. En otros casos, ha sido relevante la familiaridad tecnológica de las y los participantes. Por otra parte, también existen estudios que no han encontrado diferencias en test que miden conocimientos o habilidades comunicacionales e incluso medidas cognitivas que han logrado ser más sensibles en un





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

formato informatizado (Brandwayn et al., 2020; Noyes & Garland, 2008; Yao, 2020).

López y Labbé (2018) enfatizan que “no es posible asumir equivalencia sin evidencia”, por lo que esta debe ser previamente demostrada por los desarrolladores de test, a fin de garantizar una adecuada interpretación de los resultados. En la misma línea, las directrices internacionales sobre el adecuado y ético uso de test señalan que es responsabilidad de los desarrolladores de pruebas la documentación y difusión de las propiedades psicométricas de un test informatizado, así como también de generar evidencia clara y fundamentada sobre la equivalencia de la medida cuando se generan versiones en distintos dispositivos o modalidades. Concretamente, estas directrices refieren que se debe demostrar que las dos versiones tienen confiabilidades comparables y correlacionan entre sí a un nivel esperable además de producir medias y desviaciones estándar calibradas que generen puntuaciones comparables. Cabe destacar que los estudios de equivalencia y el desarrollo de datos normativos debe realizarse con participantes que contesten el instrumento en las mismas condiciones que experimentará la población para quien está destinado el test (Brandwayn et al., 2020; ITC, 2005; López & Labbé, 2018; Lozzia et al., 2009)

Dadey et al. (2018) señalan que la evidencia de comparabilidad debe generarse cada vez que un test se desarrolla en diferentes dispositivos (por ejemplo, computador de escritorio, portátil o Tablet) aludiendo a que lo relevante es demostrar que las puntuaciones resultantes de la administración en diferentes dispositivos son comparables, más que los dispositivos en sí mismos. Estos autores además señalan que, en el desarrollo de test en distintos dispositivos, es importante monitorear periódicamente la comparabilidad y mitigar todos aquellos aspectos que puedan afectarla.

Cabe destacar que uno de los problemas relevantes en la evaluación de la equivalencia entre formatos de un test radica en el proceso cognitivo o motriz que debe realizar el/la evaluado/a para contestar, ya que a pesar de la gran similitud que pueda existir entre ambos, en su examen psicométrico pueden aparecer como no equivalentes (Brandwayn et al., 2020; Lozzia et al., 2009; Noyes & Garland, 2008)

El presente estudio tuvo como principal objetivo recoger evidencia sobre la comparabilidad de las puntuaciones obtenidas en Hal2.0, batería de Propensión a la Accidentabilidad, en una





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

muestra de participantes que la contestaron de manera contrabalanceada en versión Tablet y en Computador (versión web). Hal2.0 es un instrumento informatizado en modalidad de aplicación controlada, es decir, que no hay control personal directo en el momento de la administración y el test solo puede ser contestado por aquellos/as evaluados/as que disponen de una clave de acceso (Bartram, 2006; Lozzia et al., 2009). Hal2.0 es un test desarrollado a la base de un modelo comprensivo de la propensión a la accidentabilidad que incluye la evaluación de las dimensiones de personalidad y cognición.

La Batería Hal2.0 estima la propensión a la accidentabilidad a partir de las siguientes 5 pruebas:

- *Prueba de Atención Dividida*: esta es una prueba de dificultad creciente, conforme se realizan cada uno de los 10 intentos que tiene la prueba. En cada uno de ellos la persona evaluada debe atender dos estímulos a la vez. En el lado izquierdo de la pantalla aparecen números del 1 al 9 (uno a la vez), frente a los cuales el evaluado debe determinar si son pares o impares, teniendo 3 segundos como máximo para dar una respuesta, utilizando la mano izquierda para ello. En la versión Tablet, la selección se realiza con un movimiento con el dedo índice (hacia arriba o hacia abajo dependiendo de si es par o impar); mientras que en la versión web debe marcar en el teclado W o S. En paralelo, en el lado derecho de la pantalla aparecen una serie de esferas de dos colores (rojas y azules) que se mueven en una pantalla. Luego de unos segundos las esferas rojas cambian a color azul y el/la evaluado/a debe seleccionar aquellas que eran de color rojo al inicio del intento.
- *Prueba de Atención al Detalle*: esta es una prueba en la que los/las evaluados/as deben determinar si dos pares de estímulos son iguales o distintos, para lo cual cuentan con tiempo límite de 1 minuto. Estos estímulos son verbales o numéricos.
- *Prueba de Atención Focalizada*: esta es una prueba en la que los/las evaluados/as deben determinar si una serie de 120 estímulos dispuestos en filas son iguales o distintos a 3 estímulos indicados en la parte superior de la pantalla. Todos los estímulos son similares en color y forma.





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

- *Prueba de Inteligencia Fluida*: esta prueba consiste en completar series de patrones o matrices que están incompletas, seleccionando el estímulo que se sigue de la regla lógica presentada.
- *Escalas de Personalidad*: el instrumento cuenta con escalas de autoreporte de rasgos de personalidad, dentro de los cuales están los rasgos de responsabilidad, respeto a las normas y propensión al riesgo.

Las hipótesis planteadas en este estudio fueron las siguientes:

H₁: Las medidas obtenidas en la prueba de atención dividida (aciertos y errores en clasificación de números, y aciertos y comisiones en la selección de esferas) a través de la aplicación en Tablet o PC será equivalente entre ambos dispositivos, y no se encontrarán diferencias estadísticamente significativas en el promedio.

H₂: La cantidad de aciertos en la prueba de atención al detalle obtenidas a través de la administración en Tablet o PC será equivalente entre ambos dispositivos, y no se encontrarán diferencias estadísticamente significativas en el promedio.

H₃: La cantidad de aciertos en la prueba de atención focalizada a través de la aplicación en Tablet o PC será equivalente entre ambos dispositivos, y no se encontrarán diferencias estadísticamente significativas en el promedio.

H₄: La cantidad de aciertos en la prueba de inteligencia fluida a través de la administración en Tablet o PC será equivalente entre ambos dispositivos, y no se encontrarán diferencias estadísticamente significativas en el promedio.

H₅: El promedio de las escalas de personalidad obtenidas a través de la aplicación en Tablet o PC será equivalente entre ambos dispositivos.

H₆: El puntaje total de propensión a la accidentabilidad será equivalente entre ambos dispositivos, y no se encontrarán diferencias estadísticamente significativas en el promedio.





Método

Participantes

La muestra fue de tipo no probabilística y de autoselección, compuesta por un total de 20 participantes, dentro de los cuales el 85% declararon ser de sexo femenino. La edad de las y los participantes fluctuó entre 19 y 45 años ($M=29$, $DS=7.8$). En relación al nivel educacional el 20% contaba con Estudios de Posgrado completos, el 35% con Enseñanza Superior Completa, 35% Enseñanza Media Completa, 5% Enseñanza Superior Técnico Profesional y 5% Enseñanza Media Técnico Profesional.

Los criterios de inclusión fueron ser mayor de 18 años de edad, contar con conexión a internet y con los dos dispositivos: Tablet y Computador.

Instrumentos

– **Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0**, en formato Tablet y Computador (Pizarro y Rosas, 2018). Esta batería, basada en un modelo multidimensional de propensión a la accidentalidad, está compuesta por 5 pruebas, con las cuales se evalúan diversos constructos asociados tanto a la dimensión cognitiva como de la personalidad.

La batería Hal2.0 fue originalmente desarrollada para ser aplicada en formato Tablet, en modalidad de administración supervisada (Bartram, 2006; Lozzia et al., 2009), es decir, en presencia de un/a evaluador/a quien guía el proceso de evaluación. Para la presente investigación se desarrolló su adaptación a formato computador (PC), para ser contestada de manera autónoma, es decir, sin asistencia de un/a evaluador/a (modalidad de administración controlada según Bartram, 2006 y Lozzia et al., 2009).

Cabe destacar que se solicitó a los/as participantes contestar la batería en los dispositivos en un orden particular, de manera contrabalanceada en la muestra por lo que la mitad de los/a participantes (10 personas) primero rindieron la versión Tablet del instrumento, y luego la versión en Computador (versión web). La mitad restante siguió la secuencia inversa.

– **Cuestionario de Prepiloteo Hal2.0**: Instrumento tipo cuestionario de autoreporte en formato online desarrollado a través de la herramienta de formularios de Google. Este cuestionario tuvo





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

como objetivo indagar la percepción de los/as participantes acerca de la experiencia de contestar la batería en sus dos formatos y debía ser contestado al momento de terminar el proceso de evaluación. El cuestionario estuvo compuesto por doce preguntas en formato de respuesta abierta y cerrada, que incluían datos de identificación, solicitud de las fechas en que las y los participantes completaron las evaluaciones, reporte de las características de los dispositivos que utilizaron (marca, modelo, pulgadas, navegador utilizado, etc.) e impresiones en torno a la administración en ambos dispositivos, eventuales dificultades y su percepción sobre la presencia o no de diferencias entre ambos.

Procedimiento

A fin de realizar el reclutamiento de la muestra se envió una invitación online junto con el consentimiento informado a través de las redes sociales de CEDETi UC, entre los meses de junio y noviembre de 2021. El consentimiento informado incluía datos de identificación y referencias al propósito de la investigación, en qué consistía, duración y el carácter voluntario de la participación en esta, señalando que no se corría ningún riesgo por ello. Además, se incluyó información sobre el manejo de los datos y la confidencialidad. Por otra parte, el consentimiento informaba que al finalizar las actividades se obtendría una retribución económica por la participación.

A los/as participantes que contestaron la invitación y expresaron su deseo de participar en el consentimiento informado se les contactó vía correo electrónico y se les envió un instructivo sobre lo que debían realizar. Este instructivo señalaba que debía realizar la misma evaluación en dos dispositivos (Tablet y PC), señalándoles por cuál debían partir y solicitando que contestaran en el segundo dispositivo justo una semana después y ojalá dentro del mismo bloque horario en que lo hicieron con el primero. Luego de esto cada participante recibió las credenciales (para Tablet) o en link (para PC) en el que debían contestar la evaluación.

Una vez que los/as participantes contestaron la evaluación en los dos dispositivos y previo chequeo de haber cumplido con los requisitos solicitados, se les envió vía correo electrónico el Cuestionario de Pre-pilotaje de Hal2.0, y luego de que lo contestaran recibieron la retribución por su participación en este estudio.





Análisis de datos

Para determinar si el formato de aplicación tiene algún impacto en el rendimiento en la prueba, se buscó despejar el efecto de la práctica del de formato de evaluación, ya que el segundo intento es realizado habiendo ya conocido y ensayado las pruebas en el primer intento (en la primera semana de evaluación). Para ello, los puntajes de las distintas subpruebas fueron analizados visualmente, mediante un gráfico que muestra el cambio en el puntaje de cada persona evaluada según el formato de la prueba rendida, y el intento (si es que era el primer o segundo intento). Luego se presenta un gráfico en que se muestra el puntaje promedio de cada indicador (en el eje y), distinguiendo el formato de aplicación (en colores) y el intento (en el eje x). Una línea que une el intento 1 y el intento 2 busca dar cuenta del cambio observado por efecto del aprendizaje, mientras que la distancia entre las líneas da cuenta del efecto del formato.

Para estimar el efecto aprendizaje y el efecto del formato se utilizó regresión múltiple, a través de modelos en donde la variable dependiente es el puntaje del indicador, y las variables independientes son el intento (1 o 2) y el formato (Tablet o Web). Al tratarse de variables categóricas binarias, el efecto se entiende como la diferencia de un intento o formato respecto a una categoría de referencia. En este caso se trata de la diferencia promedio entre el intento 2 respecto al intento 1, y la diferencia promedio del formato Web respecto al formato Tablet.

Por último, la Batería Hal2.0 entrega un puntaje compuesto, a través de diversos puntajes ponderados. Se compararon los promedios de los puntajes utilizados para el cálculo del puntaje total utilizando Pruebas t, junto con la comparación del promedio final de la prueba. Para el análisis de los datos se utilizó el software estadístico R (Core Team, 2021).





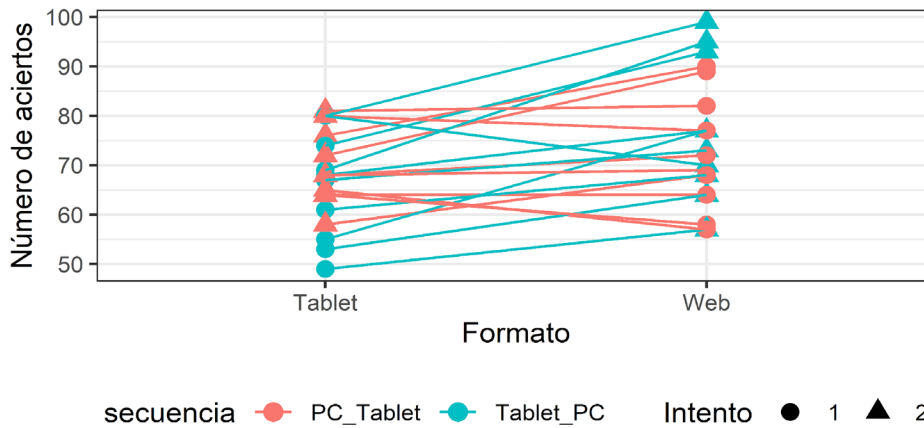
Resultados

Efecto del formato:

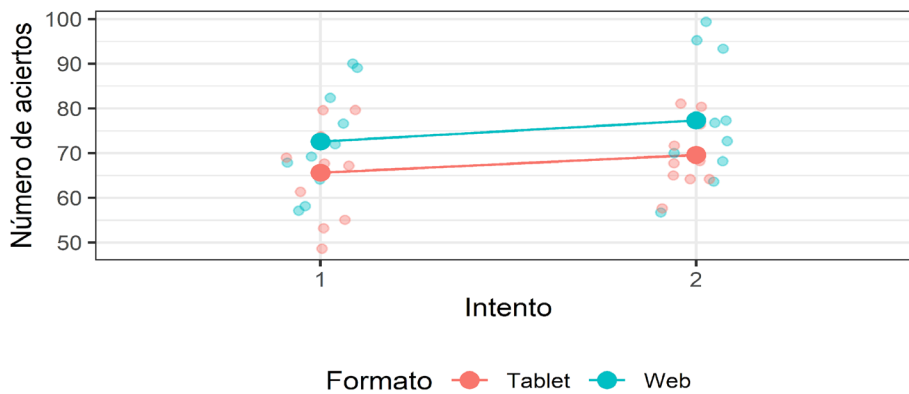
Hal2.0. Análisis de las pruebas de la dimensión cognitiva:

Atención Dividida: Aciertos en la clasificación de números

Aciertos intrasujeto en clasificación de números
Se distinguen intento (1 y 2) y formato (Web, Tablet)



Promedio de aciertos en la clasificación de números
Se distinguen los intentos (1 y 2) y el formato (Tablet, Web)





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

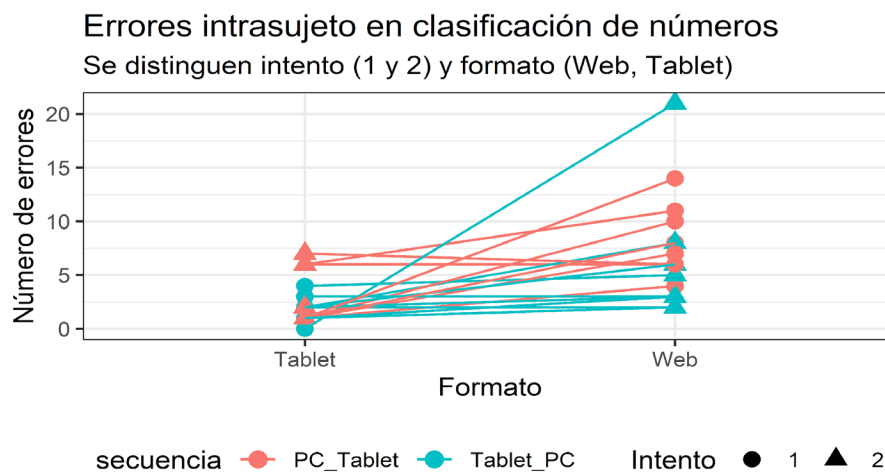
Tabla 1

Atención Dividida. Aciertos en la Clasificación de Números

	β	$ES \beta$	95% CI	p
Intercepto	65.425	3.055	59, 72	0.000
Intento 2	4.350	3.528	-2.8, 11	0.225
Versión Web	7.350	3.528	0.20, 14	0.044

La diferencia entre el formato web y el formato Tablet es de 7.35 aciertos y es estadísticamente significativa. No se observa un efecto de aprendizaje, ya que no hay un incremento en el puntaje promedio en el intento 2 respecto al intento 1 (Ver Tabla 1).

Atención Dividida: Errores en la clasificación de números





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

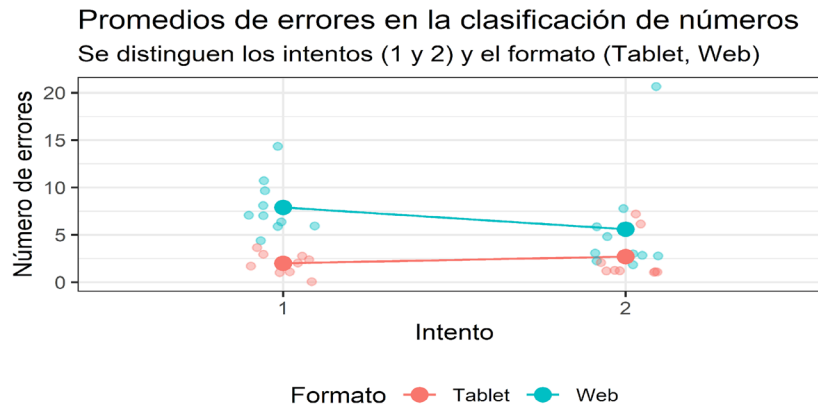


Tabla 2

Atención Dividida. Errores en la Clasificación de Números

	β	ES β	95% CI	p
Intercepto	2.75	0.974	0.78, 4.7	0.008
Intento 2	-0.80	1.125	-3.1, 1.5	0.481
Versión Web	4.40	1.125	2.1, 6.7	0.000

En promedio, quienes rinden la versión Web tienen 4.4 errores más de clasificación de números, diferencia que es estadísticamente significativa. Adicionalmente, no se observa un efecto de aprendizaje, ya que no hay un incremento en el puntaje promedio en el intento 2 respecto al intento 1 (Ver Tabla 2).





Atención Dividida: Aciertos en la selección de círculos

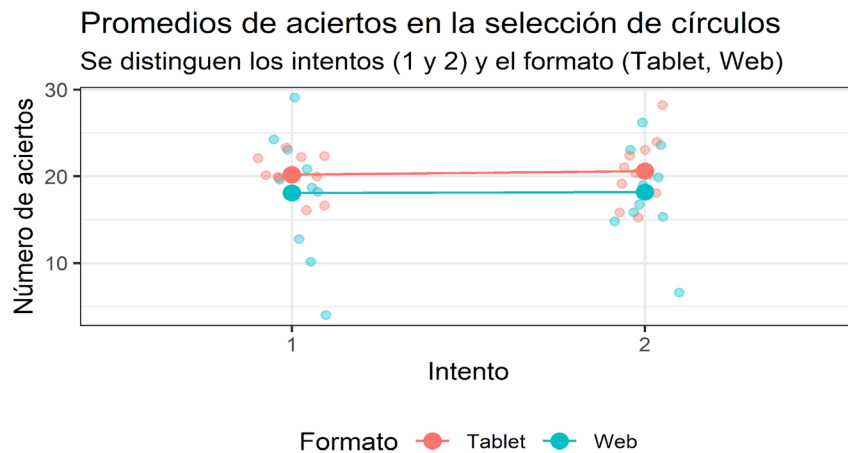
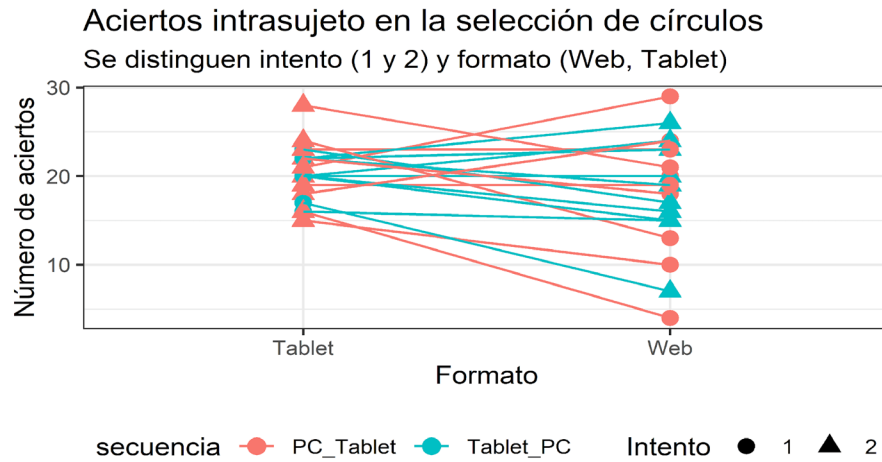


Tabla 3

Atención Dividida. Aciertos en la Selección de Círculos

	β	<i>ES</i> β	95% <i>CI</i>	<i>p</i>
Intercepto	20.275	1.378	17, 23	0.000
Intento 2	0.250	1.591	-3.0, 3.5	0.876
Versión Web	-2.250	1.591	-5.5, 1.0	0.166





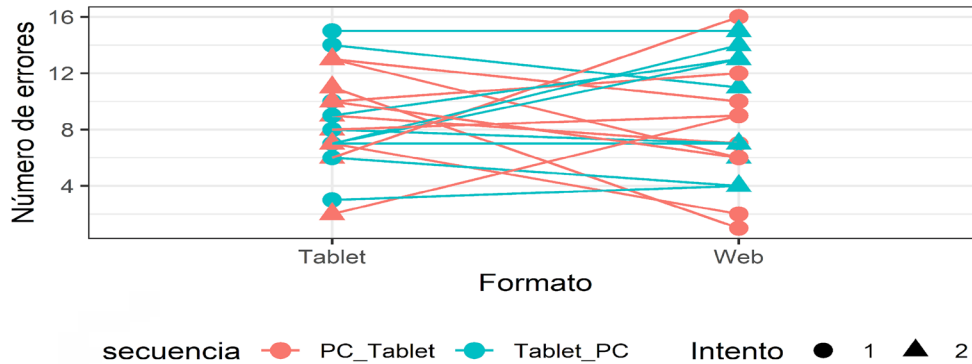
Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

Según estos resultados en la versión Web habría 2.2 aciertos menos que en la versión Tablet, diferencia que no es estadísticamente significativa. No se observa un efecto de aprendizaje, ya que no hay un incremento en el puntaje promedio en el intento 2 respecto al intento 1 (Ver Tabla 3).

Atención Dividida: Errores en la selección de pelotas

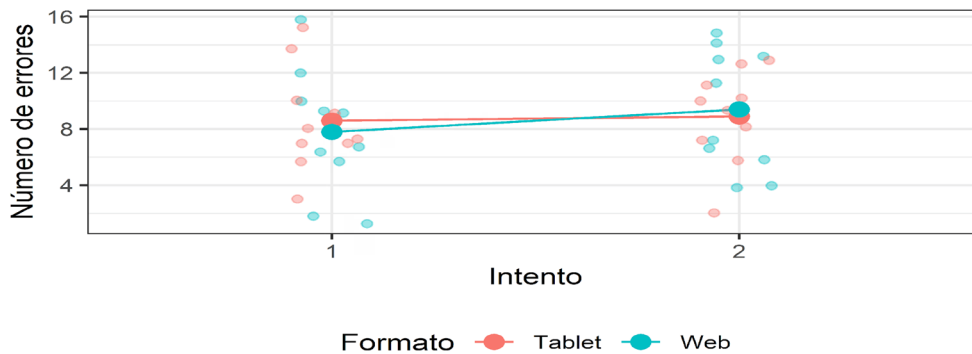
Errores intrasujeto en la selección de círculos

Se distinguen intento (1 y 2) y formato (Web, Tablet)



Promedios de errores en la selección de círculos

Se distinguen los intentos (1 y 2) y el formato (Tablet, Web)





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

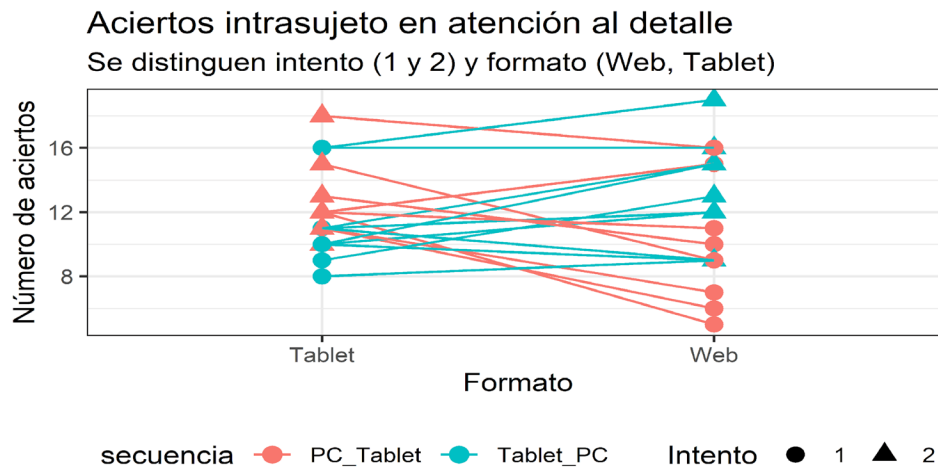
Tabla 4

Atención Dividida. Errores en la Selección de Círculos

	β	$ES \beta$	95% CI	p
Intercepto	8.275	1.071	6.1, 10	0.000
Intento 2	0.950	1.236	-1.6, 3.5	0.447
Versión Web	-0.150	1.236	-2.7, 2.4	0.904

Estos resultados (Ver Tabla 4) indican que no habría diferencia en el formato de aplicación de la prueba para efecto de determinar la cantidad de errores en la selección de círculos en la prueba de atención dividida. Además, no se observa un efecto de aprendizaje, ya que no hay un incremento en el puntaje promedio en el intento 2 respecto al intento 1. Los resultados obtenidos para las tareas de atención dividida permiten aceptar parcialmente la H_1 , de este estudio.

Atención al Detalle: Aciertos





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

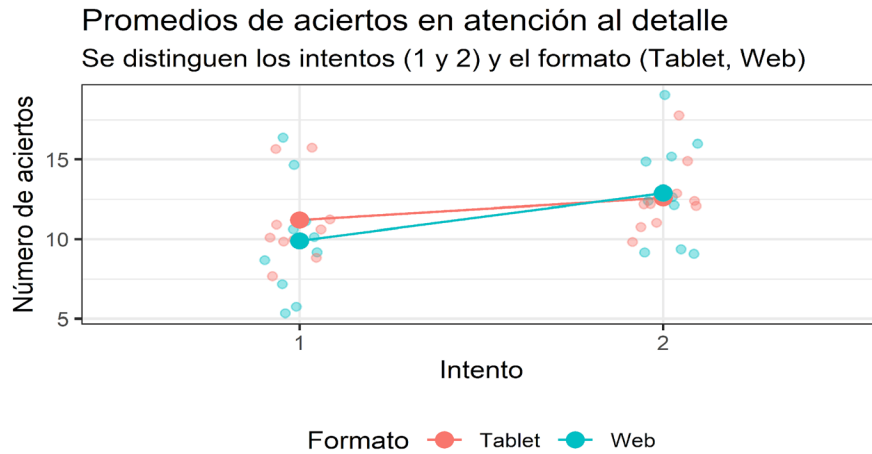


Tabla 5

Atención al Detalle. Aciertos

	β	$ES \beta$	95% CI	p
Intercepto	10.8	0.828	9.1, 12	0.000
Intento 2	2.2	0.956	0.26, 4.1	0.027
Versión Web	-0.5	0.956	-2.4, 1.4	0.604

Según estos resultados no habría diferencia en el formato de aplicación de la prueba para efecto de determinar la cantidad de aciertos en la prueba de atención al detalle, por lo que se acepta la H_2 de este estudio. Por otra parte, se aprecia una asociación significativa entre el intento y el puntaje, que evidencia un incremento de 2.2 puntos en el puntaje en el segundo intento, respecto del primero. Ello permite asumir que hubo un aprendizaje para esta prueba, asumiéndose que haber tenido un primer intento facilita la ejecución del segundo intento de esta prueba (Ver Tabla 5).



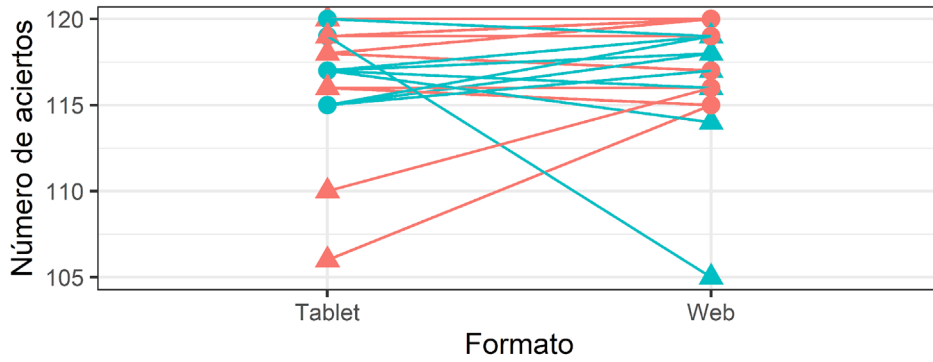


Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

Atención Focalizada: Aciertos

Acertos intrasujeto en atención focalizada

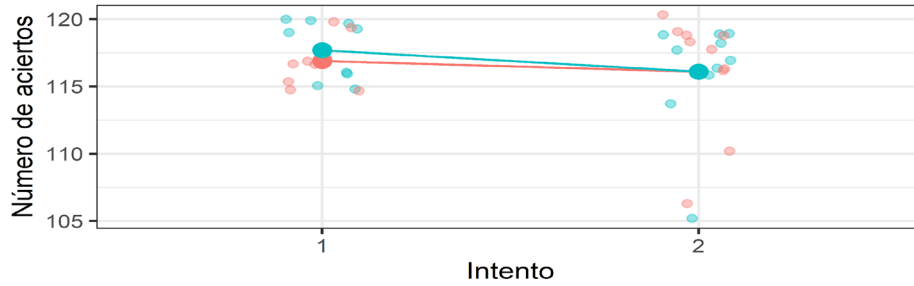
Se distinguen intento (1 y 2) y formato (Web, Tablet)



secuencia ● PC_Tablet ● Tablet_PC Intento ● 1 ▲ 2

Promedios de aciertos en prueba de atención focalizada

Se distinguen los intentos (1 y 2) y el formato (Tablet, Web)



Formato ● Tablet ● Web





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

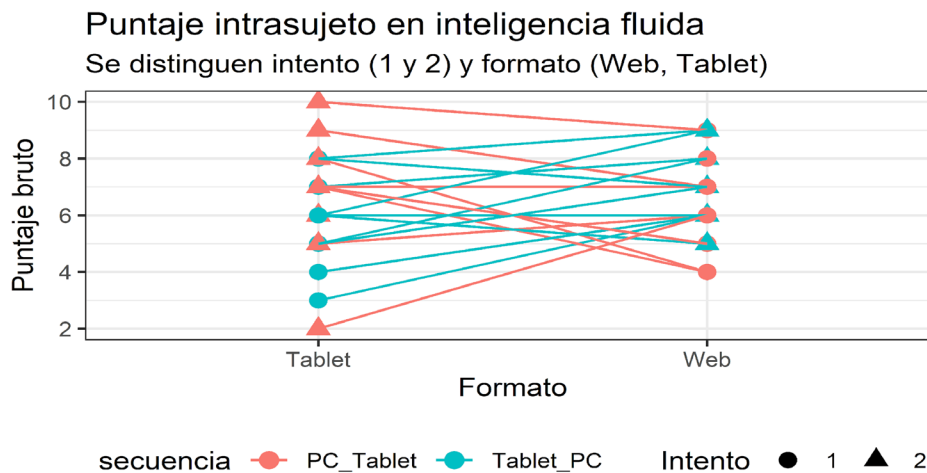
Tabla 6

Atención al Focalizada. Aciertos

	β	<i>ES</i> β	95% <i>CI</i>	<i>p</i>
Intercepto	117.1	0.917	115, 119	0.000
Intento 2	-1.2	1.058	-3.3, 0.94	0.264
Versión Web	0.4	1.058	-1.7, 2.5	0.708

Según estos resultados (Ver Tabla 6) no habría diferencia en el formato de aplicación para efecto de determinar la cantidad de aciertos en la prueba de atención focalizada, lo que permite aceptar la H₃ propuesta para esta investigación. Además, no se observa un efecto de aprendizaje, ya que no hay un incremento en el puntaje promedio en el intento 2 respecto al intento 1.

Inteligencia Fluida: Aciertos





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

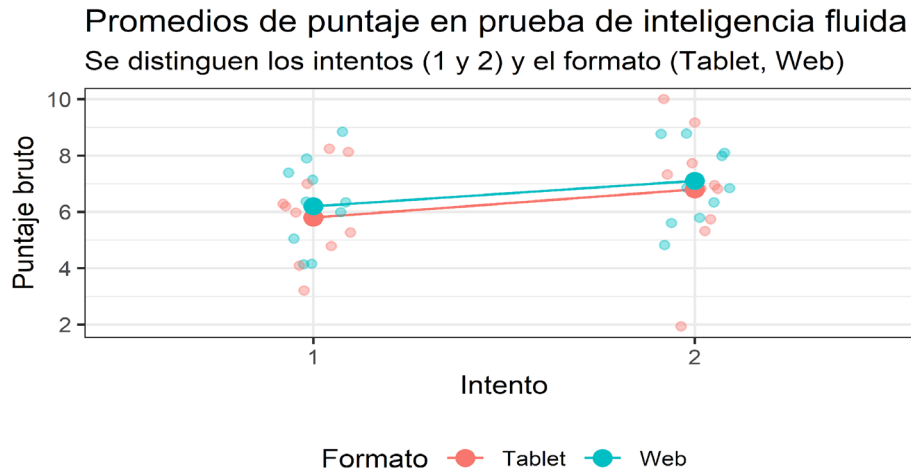


Tabla 7

Inteligencia Fluida. Aciertos

	β	$ES \beta$	95% CI	p
Intercepto	5.825	0.467	4.9, 6.8	0.000
Intento 2	0.950	0.540	-0.14, 2.0	0.087
Versión Web	0.350	0.540	-0.74, 1.4	0.521

Según estos resultados no habría diferencia en el formato de aplicación de la prueba para efecto de determinar el puntaje de la prueba de inteligencia fluida, por lo que se acepta la H_4 de este estudio. Además, si bien se observa a nivel gráfico un efecto de aprendizaje, al tener el intento 2 puntajes promedio más altos en promedio (de una respuesta correcta más, aproximadamente), este no alcanza a ser estadísticamente significativo. Esto puede en gran medida deberse por el bajo tamaño muestral (Ver Tabla 7).





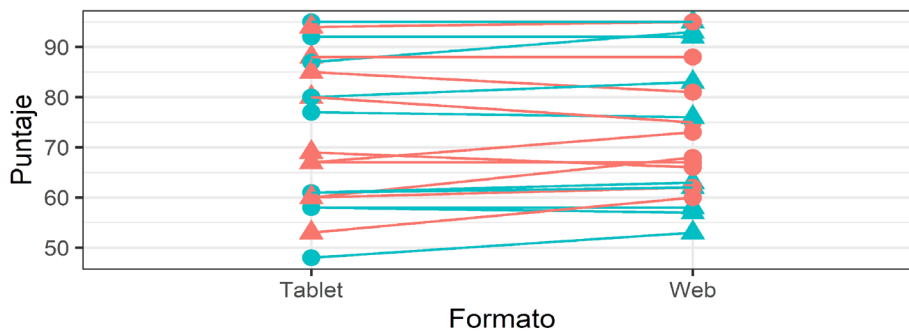
Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

Hal2.0. Análisis de rasgos en la dimensión de personalidad:

Personalidad: Responsabilidad

Puntaje intrasujeto en escala de responsabilidad

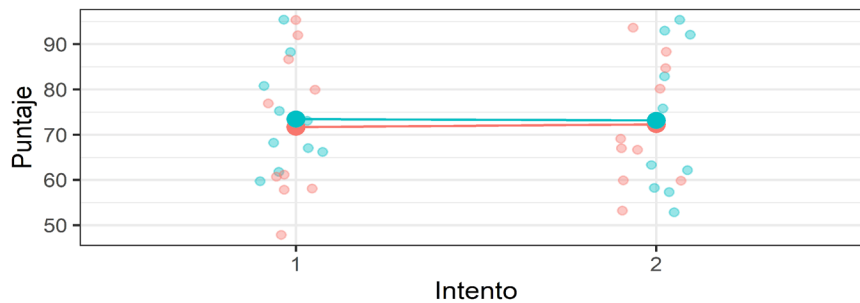
Se distinguen intento (1 y 2) y formato (Web, Tablet)



secuencia ● 1 ▲ 2

Promedios de puntaje en escala de responsabilidad

Se distinguen los intentos (1 y 2) y el formato (Tablet, Web)



Formato ● Tablet ▲ Web





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

Tabla 8

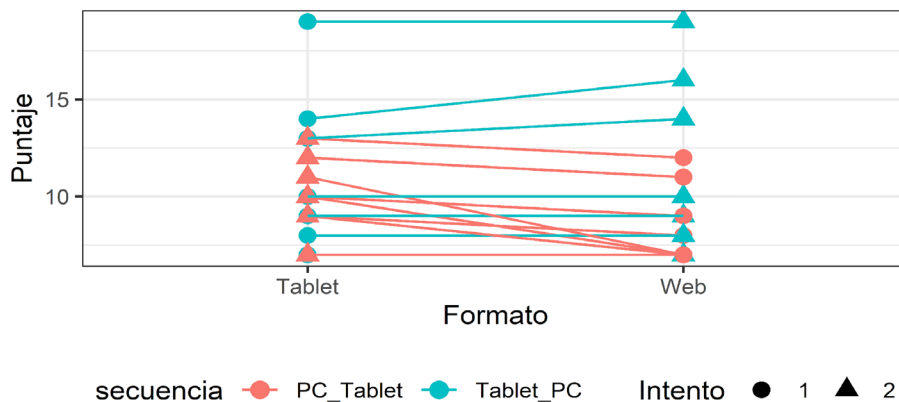
Personalidad. Responsabilidad

	β	$ES \beta$	95% CI	p
Intercepto	71.925	3.967	64, 80	0.000
Intento 2	0.150	4.581	-9.1, 9.4	0.974
Versión Web	1.350	4.581	-7.9, 11	0.770

Según estos resultados no habría diferencia en el formato de aplicación de la prueba para efecto de determinar el puntaje en la escala de responsabilidad. Además, no se observa una desviación entre el intento 1 y 2, lo cual es esperable dado que no se espera que pueda existir un efecto de aprendizaje en este tipo de escalas, y se espera una alta consistencia entre las respuestas dadas a distintos momentos. Los análisis gráficos dan cuenta de la estabilidad de las puntuaciones (Ver Tabla 8).

Personalidad: Respeto a las Normas

Puntaje intrasujeto en escala de respeto a las normas
Se distinguen intento (1 y 2) y formato (Web, Tablet)





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

Promedios de puntaje en escala de respeto a las normas
Se distinguen los intentos (1 y 2) y el formato (Tablet, Web)

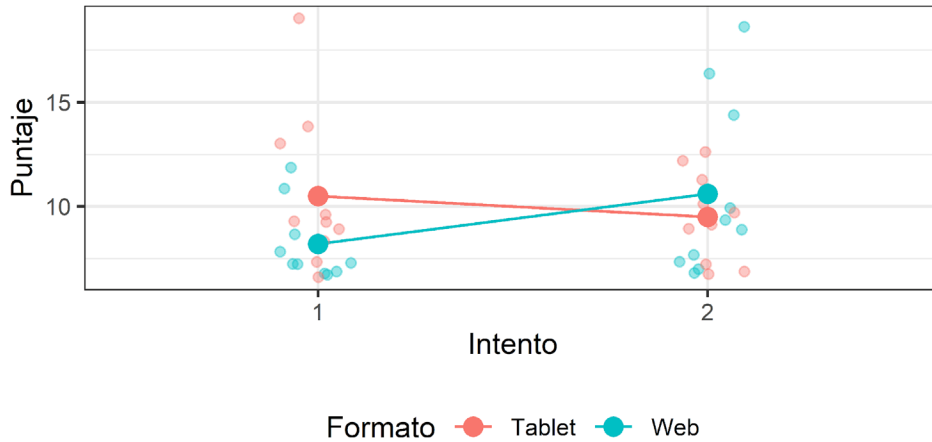


Tabla 9

Personalidad. Respeto a la Normas

Table with 5 columns: Parameter, beta, ES beta, 95% CI, p. Rows: Intercepto, Intento 2, Versión Web.

Según estos resultados (Tabla 9) no habría diferencia en el formato de aplicación de la prueba para efecto de determinar el puntaje en la escala de respeto a las normas. Al igual que para la escala anterior, no se observa una desviación entre el intento 1 y 2. Los análisis gráficos dan cuenta de la estabilidad de las puntuaciones intrasujeto, lo cual es esperable. En el entendido de que el autoreporte de características de personalidad no debiera verse afectado por el tipo de dispositivo utilizado.

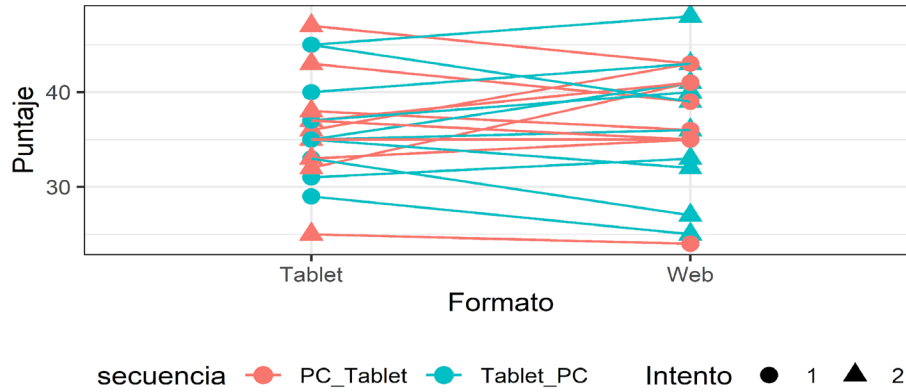




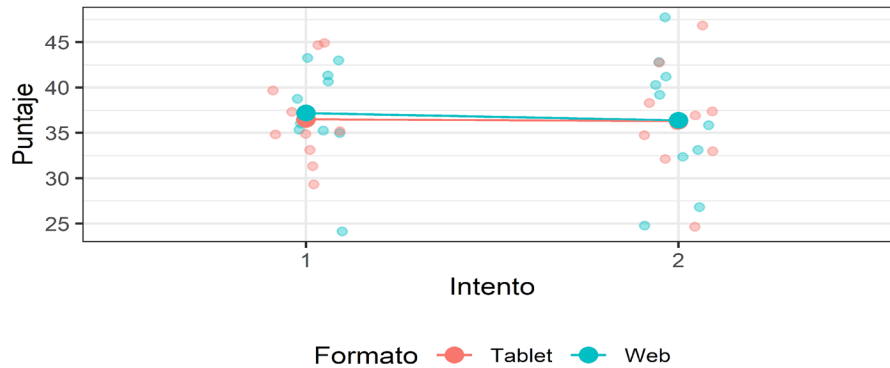
Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

Personalidad: Propensión al Riesgo

Puntaje intrasujeto en escala de propensión al riesgo
Se distinguen intento (1 y 2) y formato (Web, Tablet)



Promedios de puntaje en escala de propensión al riesgo
Se distinguen los intentos (1 y 2) y el formato (Tablet, Web)





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

Tabla 10

Personalidad. Propensión al Riesgo

	β	ES β	95% CI	p
Intercepto	36.65	1.650	33, 40	0.000
Intento 2	-0.50	1.906	-4.4, 3.4	0.794
Versión Web	0.40	1.906	-3.5, 4.3	0.835

Según estos resultados no habría diferencia en el formato de aplicación de la prueba para efecto de determinar el puntaje en la escala de propensión al riesgo. Al igual que para la escala anterior, no se observa una desviación entre el intento 1 y 2, mostrando estabilidad en las puntuaciones intrasujeto (Ver Tabla 10).

Los resultados obtenidos en la escala de personalidad permiten aceptar la H₅ de esta investigación.

Hal2.0. Puntaje Total de Propensión a la Accidentabilidad

La tabla que se presenta a continuación incluye los promedios de puntajes Z para la ejecución en formato Tablet y en la versión Web de los indicadores utilizados para el cálculo del puntaje total. Al final de la Tabla se muestra la diferencia en los puntajes totales para ambos formatos.

Tabla 11

Propensión a la Accidentabilidad, Puntaje Total

	Tablet (N=20)	PC.Web (N=20)	p value
Atención focalizada			0.709
Mean (SD)	0.354 (0.536)	0.290 (0.534)	
Range	-0.204 - 2.025	-0.204 - 2.184	





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

Comisiones números			< 0.001
Mean (SD)	-0.641 (0.148)	-0.308 (0.348)	
Range	-0.819 - -0.289	-0.667 - 0.771	
Comisiones pelotas			0.904
Mean (SD)	-0.247 (0.644)	-0.276 (0.819)	
Range	-1.526 - 0.937	-1.716 - 1.127	
Atención al detalle			0.864
Mean (SD)	-1.984 (0.798)	-1.926 (1.274)	
Range	-3.723 - -0.245	-3.723 - -0.245	
Inteligencia fluida			0.744
Mean (SD)	-0.794 (0.776)	-0.879 (0.856)	
Range	-1.560 - 0.993	-1.560 - 0.993	
Respeto a las Normas			0.459
Mean (SD)	-0.284 (0.657)	-0.444 (0.695)	
Range	-0.958 - 1.556	-0.958 - 1.327	
Puntaje Total			0.618
Mean (SD)	48.482 (3.519)	47.858 (4.303)	
Range	43.836 - 58.749	42.975 - 59.808	

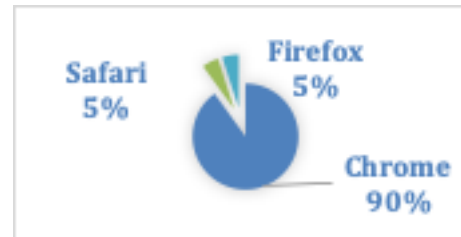
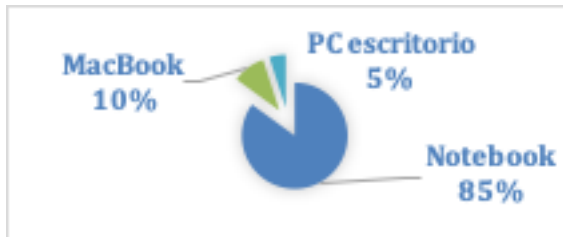
Estos resultados muestran que el único puntaje en donde existe una diferencia estadísticamente significativa en el promedio, entre los dos formatos de aplicación, es el de clasificación de números en pares e impares, de la prueba de Atención Dividida. A pesar de esto, dicha diferencia no impacta en los puntajes totales. Adicionalmente, se calculó el promedio y la desviación estándar del valor absoluto de la diferencia de puntaje T totales en los distintos formatos de aplicación. El resultado de aquella operación muestra una diferencia promedio de 2.82 puntos, y una desviación estándar de 2.36. Estos datos permiten concluir que no hay razones para suponer grandes variaciones en la obtención de puntajes entre ambas versiones del instrumento, lo que permite aceptar la H₆ de este estudio (Ver Tabla 11).



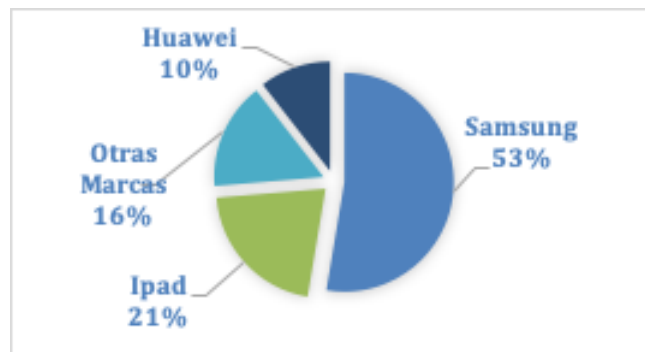


Resultados Cuestionario de Prepilotaaje Hal2.0

Al ser consultados sobre el tipo de computador utilizado, las y los participantes señalan en su mayoría haber usado un Notebook y Chrome como navegador.



En relación a la Tablet, la mayoría refiere haber utilizado una marca Samsung y las pulgadas varían entre 7 y 10.4.



Las respuestas a la pregunta abierta sobre impresiones o percepción de la evaluación con ambos dispositivos se organizan en tres categorías:

- 1) Respuestas que refieren mayor facilidad y dinamismo al contestar en Tablet, por ejemplo:

“En la tablet resulta un poco más fácil resolver las pruebas por la función táctil”

- 2) Respuestas que refieren mayor facilidad y comodidad al contestar en Computador, por ejemplo:

“Al poder hacer click y pulsar teclas se me hizo más cómodo realizar los ejercicios”.





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

- 3) Respuestas que aluden a que la actividad les resultó entretenida y/o desafiante (por ejemplo:

“Resultó interesante, a ratos parecía más una prueba de inteligencia que de propensión a la accidentabilidad”

“En cuanto a los contenidos me sorprendió el tipo de ejercicios para realizar requerían de lógica y concentración. Me interesó mucho la propuesta y su dinámica”

Discusión

De acuerdo con los lineamientos contenidos en las directrices internacionales sobre el adecuado y ético uso de test es necesario generar evidencia robusta sobre la equivalencia de la medida cuando se generan versiones de un mismo instrumento en distintos dispositivos o formatos. Para esto es necesario diseñar estudios que logren demostrar que los puntajes obtenidos en las dos versiones evidencian medias y desviaciones estándar comparables y que, correlacionan entre sí, en muestras de sujetos que contesten el test en las mismas condiciones que lo harán las personas para quienes fue diseñado (ITC, 2005; López & Labbé, 2018). En este contexto el presente trabajo buscó generar evidencia sobre la comparabilidad de las puntuaciones obtenidas en la Batería de Propensión a la Accidentabilidad Hal2.0, en una muestra de sujetos que de manera contrabalanceada, la contestaron en versión Tablet y en Computador (versión web). Cabe destacar que la mayoría de las y los participantes coincidieron en el tipo de Computador, navegador y marca de Tablet utilizada.

En la dimensión cognitiva que se evalúa con la batería Hal2.0, los resultados para la medición del constructo de *Atención Dividida* evidencian diferencias estadísticamente significativas entre las puntuaciones obtenidas por un mismo sujeto contestando en distintos dispositivos observándose una mayor cantidad, tanto de aciertos como de errores, en la versión web para la tarea de clasificar números en pares e impares, sin embargo, esto no impacta en el puntaje total de la batería. Una posible explicación a esto podría ser que la tarea de presionar teclas resulta más sencilla que la de hacer un movimiento con el dedo hacia arriba o abajo en una Tablet, lo cual





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

permite a los evaluados generar un número mayor de respuestas.

En la medida de *Atención al Detalle* los resultados indican que no habría diferencia en las puntuaciones que puedan vincularse al formato de aplicación de la prueba. Se aprecia además una asociación significativa entre el intento y el puntaje, estimador que da cuenta del incremento presentado por el puntaje en el segundo intento, respecto al primer intento, lo que da cuenta del efecto del aprendizaje.

Los resultados para la evaluación de la *Atención Focalizada* evidencian que no existiría un efecto del formato de administración de la prueba en el desempeño del/la evaluado/a, y tampoco, efecto de aprendizaje. De la misma forma, en la medida de la Inteligencia Fluida no se observa efecto del formato, sin embargo, era esperable que las puntuaciones no difieran significativamente si se considera que la acción que hay que realizar para elegir una opción de respuesta no debería depender del formato de aplicación. El aumento en la puntuación observado tras la ejecución de la tarea en el segundo dispositivo también es esperable ya que el haber realizado recientemente la prueba puede facilitar la resolución de los ítems una vez que estos ya son conocidos.

En la dimensión de *Personalidad*, en los rasgos de Responsabilidad, Respeto a las Normas y Propensión a la Accidentabilidad, la comparación de las puntuaciones evidencia que no existen diferencias entre los dispositivos. En el segundo rasgo evaluado se aprecian mínimas diferencias que pueden atribuirse a fluctuaciones esperables en el rendimiento, sin embargo, la tendencia general es a la estabilidad de los puntajes, independiente del formato de administración.

Los resultados de este estudio evidencian la equivalencia entre los formatos Tablet y PC (versión web) de la Batería de Propensión a la Accidentabilidad Hal2.0, validando su utilización en ambos dispositivos. Se recomienda que futuras investigaciones de este tipo incluyan un mayor tamaño muestral a fin de robustecer los hallazgos.





Referencias

- Bartram, D. (2006). The internationalization of testing and new models of test delivery on the Internet. *International Journal of Testing*, 6(2), 121-131 https://www.tandfonline.com/doi/abs/10.1207/s15327574ijt0602_2
- Brandwayn, N., Restrepo, D., Macías-Bohórquez, R., López-López, W., & Acevedo-Triana, C. A. (2020). Comparación del desempeño de jóvenes en pruebas neuropsicológicas en formato lápiz y papel e informatizadas. *Diversitas: Perspectivas en Psicología*, 16(2), 365-386 <https://doi.org/10.15332/22563067.6319>
- Dadey, N., Lyons, S., & DePascale, Ch. (2018) The Comparability of Scores from Different Digital Devices: A Literature Review and Synthesis with Recommendations for Practice, *Applied Measurement in Education*, 31(1), 30-50, DOI: 10.1080/08957347.2017.1391262
- Elosua, P. (2021). Aplicación remota de Test: riesgos y recomendaciones. *Papeles del Psicólogo*, 42(1), pp. 33-37 <https://doi.org/10.23923/pap.psicol2021.2952>
- Elosua, P. (2022). Impacto de la TIC en el entorno evaluativo. Innovaciones al servicio de la mejora continua. *Papeles del Psicólogo*, 43(1), pp. 3-11. <https://doi.org/10.23923/pap.psicol.2985>
- Hernández, A., Elosua, P., Fernández-Hermida, J., & Muñiz, J. (2022). Comisión de Test: Veinticinco años velando por la calidad de los test. *Papeles Del Psicólogo*, 43(1), pp. 55-62. <https://doi.org/10.23923/pap.psicol.2978>
- International Test Commission (2005). *International Guidelines on Computer-Based and Internet Delivered Testing*. [www.intestcom.org] https://www.intestcom.org/files/guideline_computer_based_testing.pdf
- López, L., & Labbé, C. (2018). Instrumentos de recolección de datos en formato digital: No hay que asumir equivalencia sin evidencia. *Revista médica de Chile*, 146(6), 815-816. <https://dx.doi.org/10.4067/s0034-98872018000600547815>





Evidencia de equivalencia entre dispositivos de la Batería de Evaluación de la Propensión a la Accidentabilidad Hal2.0

- Lozzia, G., Abal, F., Blum, D., Aguerri, M., Galibert, M., & Attorresi, H. (2009). Test Informatizados. Nuevos desafíos prácticos y éticos para la evaluación psicológica. *SUMMA Psicológica UST*,6(1), 135 – 148 <https://dialnet.unirioja.es/servlet/articulo?codigo=3020389>
- Menéndez, L., Hierro, P., & Muñiz, J. (2008). Actitudes hacia los test informatizados aplicados por internet con formato responder hasta acertar. *Acción Psicológica*, 5(2), 25-36. <https://www.redalyc.org/articulo.oa?id=344030760004>
- Muñiz, J., & Hambleton, R. K. (1999) Evaluación psicométrica de los test informatizados. En J. Olea, V. Ponsoda & G. Prieto (Eds.), *Test informatizados: fundamentos y aplicaciones* (pp. 23-52). Pirámide.
- Noyes, J., & Garland, K. (2008) Computer- vs. paper-based tasks: Are they equivalent?, *Ergonomics*, 51(9), 1352-1375, DOI: 10.1080/00140130802170387
- Olea, J., Abad, F., & Barrada, J. (2010). Tests informatizados y otros nuevos tipos de tests. *Papeles del Psicólogo*, 31(1), 97-107 <https://www.redalyc.org/articulo.oa?id=77812441010>
- Olea, J., Ponsoda, V. y Prieto, G. (1999). *Tests informatizados Fundamentos y aplicaciones*. Madrid: Pirámide
- R Core Team (2020). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Rosas, R., & Pizarro, M. (2018). HAL2: Herramienta en formato Tablet para la evaluación de la propensión a la accidentabilidad. *CEDETi-UC Papeles de Investigación N°10*. <http://www.cedeti.cl/wp-content/uploads/2022/03/Papeles-de-investigacio%CC%81n-No10-2019.pdf>
- Yao, D. (2020). A Comparative Study of Test Takers' Performance on Computer-Based Test and Paper-Based Test across Different CEFR Levels. *English Language Teaching*, 13(1), 124-133. <https://files.eric.ed.gov/fulltext/EJ1239128.pdf>

