

Grado en PSICOLOGÍA
Trabajo Fin de Grado
Curso 2020-2021



EVALUACIÓN PSICOMÉTRICA DEL TEST DE AULA NESPLORA: APLICACIÓN

Alberto Infantes García

Trabajo Fin de Grado
Facultad de Psicología
Grado en PSICOLOGÍA
Universidad de Murcia

TÍTULO DEL TFG: Evaluación psicométrica del test AULA NESPLORA: Aplicación

Nombre del alumno/a: Alberto Infantes García

Nombre del tutor/a: María Dolores Hidalgo Montesinos

Perfil de TFG: Empírico

Curso 2020-2021

Agradecimientos

Los agradecimientos del presente Trabajo de Fin de Grado son, sobretodo, para las personas que más me han acompañado durante esta trayectoria, por un lado, mi tutora de TFG María Dolores Hidalgo Montesinos, que no solo ha dictaminado las pautas del trabajo, sino que, además, ha dedicado su tiempo a enseñarme sobre la materia y hacer correcciones muy oportunas, por lo que su apoyo y orientaciones han sido muy importantes para el desarrollo del mismo.

Por otro, Miguel Saura, por pasar conmigo la carrera y, tras ella, descubrirme el test AULA Nesplora y al cual le estoy agradecido por animarme a escoger esta línea de investigación que me ha hecho tener otra visión de la psicometría. Siempre voy a recordar el momento del café y las líneas de TFG, al final llevaba razón y es que los test informatizados hoy en día tienen mucho potencial.

Por último, agradecer a todas las personas que han estado en mi vida mientras he completado una etapa tan importante de la misma, sobretodo, a mi madre, mi hermana y mi novio que me han acompañado durante todos los momentos felices y críticos de esta experiencia.

DECLARACIÓN DE ORIGINALIDAD

El abajo firmante **DECLARA** ser el autor del Trabajo de Fin de Grado que presenta como culminación de sus estudios de Graduado en la Facultad de Psicología de la Universidad de Murcia, para ser evaluado por la Comisión Evaluadora correspondiente y **desea hacer constancia de lo siguiente:**

1. **El trabajo es original e inédito** y es producto de su contribución intelectual, bajo la tutela de uno o varios profesores de esta Facultad.
2. **Las figuras, tablas e ilustraciones** que lo acompañan al trabajo representan fielmente los hechos informados y **no han sido alteradas digitalmente.**
3. **Todos los datos y las referencias** a textos y materiales ya publicados están **debidamente identificados y referenciados en el texto y en las notas bibliográficas.**

Para dejar constancia de lo anteriormente expuesto, se firma esta declaración en Murcia, a 28 de JUNIO de 2021:

Fdo:

Índice

Resumen.....	1
Abstract	1
1. INTRODUCCIÓN	2
1.1. Marco teórico del CET-R	3
1.2. Marco teórico de AULA NESPLORA	4
2. METODOLOGÍA.....	9
2.1. Procedimiento de aplicación del CET-R	9
3. RESULTADOS.....	10
3.1. Resumen de la valoración del test	10
3.2. Validez.....	17
3.3. Fiabilidad	21
3.4. Baremación e interpretación de las puntuaciones	23
3.5. Valoración final	26
4. DISCUSIÓN.....	27
4.1. Validez convergente	27
4.2. Sensibilidad	28
4.3. Usos fuera del ámbito de la psicología clínica.....	29
5. CONCLUSIONES.....	30
5.1. Limitaciones.....	30
6. REFERENCIAS	31

Resumen

Para la evaluación psicológica es necesario contar con instrumentos con propiedades psicométricas adecuadas, por tanto, el presente trabajo trata sobre la aplicación de un cuestionario evaluativo de calidad para herramientas psicométricas (CET-R) sobre el test informatizado en formato de Realidad Virtual AULA NESPLORA para redes atencionales. En el proceso se definen ambas herramientas, sus propiedades y artículos sobre el trabajo práctico que hacen los profesionales con ellas. Los resultados indican que AULA NESPLORA es un test que marca calidad en los estándares de evaluación del CET-R pudiendo incluirse en la lista de test recomendados para su uso profesional.

Palabras clave: Evaluación neuropsicológica, Psicometría, Revisión de Calidad, Evaluación de los procesos atencionales, CPT, Realidad Virtual, AULA NESPLORA, CET-R.

Abstract

For psychological assessment it is necessary to have instruments with adequate psychometric properties, therefore, this paper deals with the application of an evaluative quality questionnaire for psychometric tools (CET-R) on the computerized test in Virtual Reality format AULA NESPLORA for attentional functions. In the process, both tools, their properties and articles about the practical work that professionals do with them are defined. The results indicate that AULA NES-PLORA is a test that marks quality in the evaluation standards of the CET-R and can be included in the list of tests recommended for professional use.

Keywords: Neuropsychological assessment, Psychometrics, Quality Review, Assessment of attentional processes, CPT, Virtual Reality, AULA NESPLORA, CET-R.

1. INTRODUCCIÓN

La tecnología hoy en día nos permite realizar nuevas medidas y técnicas de evaluación, ésta nos permite basar la medición en modelos teóricos, en la teoría del ítem, aplicar nuevas fórmulas de deseabilidad social, automatizar el registro de conductas, actualización continua de los baremos, nuevos formatos de ítems, nuevos modelos psicométricos, autoinforme e incluso presentar distintas situaciones de conflicto laboral, según Olea et al., (2010) los test informatizados podrían cubrir nuevas demandas sociales ya que la mejor estandarización de las pruebas y sus condiciones de aplicación pueden superar las exigencias de calidad.

Ésta diferencia con respecto a test clásicos se muestra a la hora de estudiar variables en estudios experimentales. Así, Rodríguez-Barranco et al., (2016) tratan de distinguir los efectos perjudiciales o positivos para la salud en exposición a trazas de metales. Para ello utilizaron diversas pruebas y encontraron que los autoinformes que completan padres y profesores no fueron válidos para enseñar asociaciones significativas. Debido a esta necesidad, surgen trabajos como el de García y González (2015) en el que tratan de recopilar las escalas específicas más conocidas y utilizadas para la evaluación de las variables atencionales o TDAH en niños y adolescentes para realizar un diagnóstico correcto y conocer el perfil de funcionamiento cognitivo.

Las ventajas que aportan los cuestionarios son la fácil aplicación y accesibilidad, sin embargo, son pruebas subjetivas, indirectas e inespecíficas y las respuestas pueden diferir de acuerdo al informante. (Herrán et al., 2014) Por otra parte, las pruebas de desempeño son objetivas y directas, aunque consumen mayor tiempo de aplicación y requieren de espacios acondicionados con equipos de computadoras o hardware no tan accesibles.

¿Qué caracteriza entonces a los test informatizados? Según Olea et al., (2010) deben cumplir dos requisitos: la disponibilidad de sus propiedades psicométricas y que los ítems se presenten y respondan en un ordenador. En el caso de AULA Nesplora (Climent y Banterla, 2016), el test se responde en un hardware de realidad virtual que consiste en unas gafas de realidad virtual, las cuales monitorizan desde un visor un software interactivo con controles, como un ordenador. Encontramos en la literatura con que los test informatizados tipo Continuous Performance Test, en adelante CPT, usualmente son conocidos

por tener una extraordinaria validez interna, pero no externa, hay revisiones con una visión crítica ante cómo los distintos paradigmas de tarea se convierten en un test, sin serlo realmente debido a que tienen una estandarización o normativo. Critican cómo los estímulos se presentan solo visual, solo auditivo o con un paradigma u otro. Además, critica la validez ecológica de las pruebas y problemas con los tiempos de respuesta (Kessels, 2018) Respecto al formato de realidad virtual, Delgado y Moreno (2012), encuentran diversas evidencias que apoyan la validez ecológica de la realidad virtual para medir procesos relacionados con la atención, con la actividad motora, o un mejor control de las situaciones ambientales.

Por todo esto, el presente trabajo busca revisar este test mediante el cuestionario de evaluación de Test (CET-R) con el objetivo de arrojar luz sobre las características de estos nuevos tipos de test y proporcionar un informe de la calidad de los mismos en línea con el trabajo de la Comisión Nacional de Test del Colegio Oficial de Psicólogos.

1.1. Marco teórico del CET-R

Históricamente los diversos oficios suelen agruparse en colegios, gremios o asociaciones que intentan encontrar apoyo para los practicantes de ese oficio a través de consensos y reglas que se estipulan con el fin de mejorar las condiciones y resultados de estos. En la comunidad de psicólogos se recogió datos sobre el gran uso y ayuda de los test como herramienta, por ende, y como naturalmente ha ido surgiendo con códigos deontológicos o convenios profesionales, de la mano del Colegio oficial de Psicólogos en España surgió también la necesidad de crear un consenso o valoración sobre la calidad o utilidad de estas herramientas, de manera que crearon un cuestionario de calidad basado en una metodología que nos permita valorar los test de una forma más sencilla a fin de elegirlos fácilmente.

Según Hernández et al. (2016) en primer lugar, se revisaron trabajos similares en otros investigadores en países pioneros como Países Bajos, Gran Bretaña, Noruega de la mano de asociaciones como la EFPA (European Federation of Psychological Associations) que propuso medidas como revisar en las herramientas los avances psicométricos y tecnológicos. En segundo lugar, se diseñó el test para la evaluación desde la Teoría Clásica de los Test creando así 3 apartados: Descripción técnica de la prueba, evaluación técnica de las características del instrumento y valoración global del instrumento. Por últi-

mo, revisaron el CET acorde con las directrices de asociaciones como la APA, AERA, NCME o ITC y tener resultados más globales y comunes a la comunidad de profesionales de la salud mental, creando así el CET-R.

Utilizando esta herramienta, la Comisión del Test del Consejo General de la Psicología en España promueve anualmente la revisión de la calidad de diferentes test publicados (Hidalgo et al., 2019). Esta revisión la organizan supervisores que estudian los informes de las aplicaciones del CET-R realizados por otros especialistas o estudiantes. Cada test recibe dos aplicaciones de CET-R de la mano de dos revisores distintos y el trabajo del supervisor es generar un informe entre los dos que ha recibido, tras ello, se manda a la editorial y a la comisión del test.

Actualmente, hay un total de 91 test evaluados (Viladrich et al., 2021) durante más de 8 revisiones en las que se han ido añadiendo más participantes y distintas sugerencias de mejora al proceso.

1.2. Marco teórico de AULA NESPLORA

Como se verá a continuación, desde el manual de AULA (Climent y Banterla, 2016) extraemos la información que contextualiza el test en estos pasos a lo largo de este punto.

Nesplora Aula surgió de la idea de adaptar el formato que tenían los test de diagnóstico de TDAH a un formato que le resultase más asequible a los niños. Esto debía de lograrse sin modificar las cualidades que tenían los test clásicos realizados en papel. En Nesplora- Technology & Behavior se había dado un gran avance en el área de la realidad virtual, y se pensó desde la organización que un escenario de realidad virtual en una clase sería muy bien aceptado por los niños. Se realizó una labor de investigación para comprobar si ya existía algún test como el que pretendían desarrollar. Se comprobó que a fecha de 2008 no existía ningún test comercial como Nesplora AULA.

En un inicio se planteó adaptar las herramientas existentes al ámbito de la realidad virtual, pero la principal barrera que se encontró fue que se necesitaba un test que fuese sólido y baremable. El fin sería dar información más realista y que carezca, en cierto sentido, de sesgos subjetivos por parte de los padres, profesores o por los mismos clínicos y objetivar, dando un valor añadido, el diagnóstico del TDAH. Esto se debe a que este trastorno tiene una crítica muy amplia con respecto a su diagnóstico. El TDAH es un trastorno

neuroconductual frecuente en la población escolar y representa uno de los principales motivos de consulta en servicios de salud mental. Su diagnóstico es complicado por la dificultad que representa la evaluación objetiva de aspectos subjetivos como la desatención o la impulsividad. Algunos profesionales defienden que está sobrediagnosticado mientras que hay otro gran número de profesionales que manifiestan también que existe un gran porcentaje de niños que carecen del diagnóstico y presentan el trastorno, esto se ve claramente en el estudio de Herrán et al., (2014) que analiza la prevalencia mostrando una alta variación a nivel mundial, entre el 2.4 y 19.8, lo que puede atribuirse a la falta de uniformidad en metodología y complejidad en los criterios evaluativos para establecer el diagnóstico de este trastorno.

En el manual se define como un trastorno del desarrollo en el que los síntomas principales son la falta de atención, la cual se entiende como los mecanismos de distribución y mantenimiento de la actividad psicológica, hiperactividad e impulsividad inapropiados que afectan al patrón de rendimiento. Existen diferentes tipos de atención, como la atención focalizada, la atención sostenida y la dividida. De estas últimas la que más relación se le ha encontrado con el TDAH es la atención sostenida.

La impulsividad como hemos indicado es otro de los ejes del TDAH. Podríamos definirla como un déficit para inhibir conductas en respuesta a demandas situacionales, la impulsividad la podríamos separar en dos tipos: la impulsividad cognitiva y la impulsividad social.

La impulsividad social sería como ejemplo la incapacidad de un niño para esperar su turno en algún juego o realizar acciones sin valorar los riesgos o asumiendo riesgos altos. Además, esta impulsividad va acompañada de una baja tolerancia a la frustración. La impulsividad cognitiva tiene como ejemplo la tendencia de un niño a responder precipitadamente a las preguntas de forma que comete errores por falta de procesamiento de la información. Por último, la sobreactividad podríamos definirla como un exceso de actividad motora o verbal de lo que cabría esperar para un niño de determinada edad. De forma general, los niños con sobreactividad suelen mantener una parte de su cuerpo en movimiento o hablan continuamente. Estos niños además tienen dificultad para mantenerse sentados en una silla por largos periodos de tiempo.

Una característica que llama la atención es que la sintomatología del TDAH aparece con tareas que requieren atención y esfuerzo, sin embargo, si la tarea presenta en sí estimulación hacia una motivación intrínseca la sintomatología del TDAH suele disminuir.

A la hora de utilizar AULA contamos con un equipo capaz de integrar en la evaluación la agitación motriz a la vez que se evalúan los distintos aspectos de la atención, cosa que es imposible de integrar con un test clásico escrito en papel.

Antes de explicar que es un CPT es conveniente definir que la atención es un proceso cognitivo básico que subyace a otras funciones cognitivas superiores. Es necesaria la atención para poder evaluar procesos como son la percepción o la memoria. Fue debido a la importancia que tiene la atención en sí como proceso por lo que se idearon los CPT, con ellos se pretendía poder medir la atención de forma aislada del resto de procesos.

Un CPT es una herramienta de evaluación informatizada que presenta estímulos organizados acorde a un paradigma para medir procesos cognitivos. La mecánica general de los CPT es que el individuo debe de reaccionar ante los estímulos presentados de forma desordenada acorde a unas instrucciones previas.

La ventaja que tienen los CPT a la hora de evaluar el TDAH es la sensibilidad que presentan ante problemas neurológicos. Otra de las ventajas que tienen los CPT es que no se ven afectados por problemas derivados del sistema motor, visual o algún tipo de dificultad del aprendizaje. Los test de este tipo tienen generalmente una gran validez a la hora de medir la atención sostenida y tienen un bajo nivel de ejecución, por lo que son asequibles tanto para población joven, adulta o anciana.

No podemos afirmar a día de hoy que exista un test que diagnostique por sí solo TDAH, pero se ha demostrado que los CPT tienen una mayor sensibilidad y especificidad. Utilizando el CPT de Conners (1995) encontró correlación entre los errores por omisión y los síntomas de inatención, así como correlación entre los errores por comisión y los síntomas de impulsividad según lo descrito en los síntomas que presentan las personas con TDAH. Al mismo tiempo se observó que el tiempo de respuesta era un indicador de sintomatología del TDAH. Estas personas con TDAH presentaban un mayor enlentecimiento al pulsar el botón frente a un estímulo, además obtenían una alta puntuación en errores por comisión, que es la capacidad para distinguir entre los estímulos diana y los estímulos

neutros. Este hecho sugiere que las personas con TDAH tienen mayor dificultad a la hora de distinguir entre los estímulos diana y los no-diana.

Inicialmente, se utilizó esta prueba para medir el número de aciertos por parte del individuo, indicando así una medida de la atención selectiva. En la actualidad los CPT tienen varias aplicaciones para obtener diferentes medidas, entre ellas la inatención, la impulsividad, la velocidad de procesamiento y la atención sostenida, dividida o alternante. Los CPT más tradicionales tienen el formato de pulsar el botón cuando aparece el estímulo X, mientras que, los más actuales tienen el formato diana o TARGET.

Nesplora Aula es un CPT Virtual y este mide la atención, impulsividad, velocidad de procesamiento, tendencia a la distracción y actividad motora durante la tarea. Estas medidas son clave debido al enfoque que tiene a la hora de elaborar un diagnóstico de TDAH con o sin hiperactividad. La evaluación va de acorde a los criterios diagnósticos que se recogen en el manual del DSM-V (American Psychiatric Association, 2013) y la CIE10 (OMS, 1992).

Con objetivo de mejorar el diagnóstico del TDAH, muchos profesionales utilizan pruebas de atención, memoria o velocidad de procesamiento, pero en la actualidad para diagnosticar basta con cumplir los criterios del DSM-V o CIE10. El principal informador serán los padres y si es posible se cotejará la información con la proporcionada por los profesores.

Algo a resaltar en este tipo de pruebas es que se reduce muchísimo el número de falsos negativos. Esto es debido a que un niño podría por ejemplo responder mal las pruebas por algún elemento externo o inoportuno, lo que resultaría para la prueba un falso positivo. Sin embargo, un niño con TDAH muy difícilmente va a poder realizar la tarea correctamente, independientemente de la situación que disponga el día de la prueba.

A continuación, se explicarán los distintos estímulos y paradigmas que utiliza Aula para medir la atención sostenida y la impulsividad.

Errores por omisión: Estos errores ocurren cuando el niño deberá de pulsar el botón reaccionando a un estímulo objetivo, pero no lo hace. El niño ha omitido el estímulo y según la literatura actual, estos errores se relacionan con la inatención. Se presentan dos posibilidades ante una persona que comete muchas omisiones. El caso en el que la persona sea sensible a situaciones monótonas y no tenga una gran capacidad de atención

(atención focalizada) o bien el caso en el que existan problemas en la gestión de la información procedente de dos canales sensoriales distintos (atención mixta).

Errores por comisión: Estos errores están más relacionados con la impulsividad. Se producen cuando la persona no debería de pulsar el botón, pero aun así lo hace. Esto indica que las personas con alta tasa de comisiones tienen una baja capacidad de inhibición de la respuesta o son demasiado impulsivas.

Tiempos de respuesta: Se mide objetivamente el tiempo en el que se lleva a cabo una tarea monótona y se valora el modo en el que se procesa, los tiempos de reacción a diferentes estímulos y la adaptación al tiempo establecido en la tarea. También se evalúa las diferencias entre algunos modelos de procesamiento, en los que se valora más bien un tiempo de procesamiento lento más que un déficit en atención.

Actividad motora: Las gafas poseen un velocímetro que permite cuantificar numéricamente la actividad motora durante la evaluación y baremarlo.

Como hemos mencionado, los estímulos se presentan de forma auditiva y visual, pero nunca ambos al mismo tiempo. La tarea en sí consta de 2 ejercicios, en el primero se muestran objetos y el estímulo diana es una manzana. En el segundo ejercicio los estímulos son números, y el estímulo diana es el número 7. Según la autora del test, estos estímulos no están seleccionados al azar, sino que responden a la población que está destinada el test. Explica que letras en niños de entre 6 o 7 años pueden no estar aprendidas correctamente y esto podría dar lugar a falsos positivos. Sin embargo, en este rango de edad los niños ya tienen los números y los objetos correctamente interiorizados.

Con respecto a los paradigmas, el primer ejercicio sigue la forma de Xno. Refiriéndonos a la tarea anteriormente descrita, el niño deberá de responder a todos los estímulos que se le presenten tanto auditiva como visualmente mientras estos no sean la palabra "manzana". Este primer ejercicio tiene una tasa de presentación de estímulos muy alta, por lo que el individuo tendrá que permanecer en un continuo estado de alerta y con la memoria de trabajo continuamente en ejecución. Con este tipo de ejercicio se podrán medir las capacidades de ejecución e inhibición.

El segundo ejercicio tiene el formato tarea X, por lo que tendrá que responder siempre que vea o escuche el número 7. En este ejercicio la tasa de estímulos es menor

que en el primero, por lo que el individuo tendrá un escenario propicio a la distracción. Esto es debido a la disminución de las demandas que se realizan a la memoria de trabajo y a un estado de alerta reducido.

Mencionar que existe un paradigma AX, que consiste en responder al estímulo di-
na X siempre que vaya precedido de otro estímulo determinado. Este paradigma se dejó
de lado en Nesplora AULA ya que requería de un mayor nivel de memoria de trabajo y de
comprensión, por lo que al tratarse de un test enfocado a niños quizás fuese demasiado
exigente para ellos. Además de esto los paradigmas de tarea X y tarea noX son más pu-
ramente atencionales, y era la atención lo que principalmente se buscaba medir con el
desarrollo de este test.

2. METODOLOGÍA

2.1. Procedimiento de aplicación del CET-R

Una vez desarrollado el marco de ambas herramientas, procederemos al desarrollo
de cómo se realizó la aplicación del CET-R.

Durante la administración se tuvieron en cuenta las distintas instrucciones que el
propio test indica para su correcta aplicación en las “OBSERVACIONES PARA TENER
EN CUENTA AL RESPONDER EL CUESTIONARIO DE EVALUACIÓN DE TESTS REVI-
SADO (CET-R)”

De manera que.

no se modificó ninguna de las preguntas del cuestionario, tratando de elegir las respues-
tas pertinentes haciendo búsqueda y formación sobre psicometría para la tarea.

En un principio, se consideró la idea de añadir NESPLORA AQUARIUM para una pers-
pectiva de trabajo sobre “Evaluación en CPTs de Realidad Virtual”. Esta quedó descarta-
da por la falta de recursos y exigencias del propio proyecto.

La documentación entregada por el fabricante junto a la herramienta consiste en una base
de contenido de libre acceso para manuales, artículos y diversas publicaciones, por lo que
se tomó la decisión de usar solo el manual como referencia.

Se realizó un escrutinio minucioso de todo el manual para el apartado de “Resultados” añadiendo en el apartado “Discusión” la información que pudiera ser interesante añadir pero que fuera extraída de otros medios.

Se argumentaron todas las calificaciones en las preguntas abiertas.

Se trató de dar respuesta a todas las preguntas, apareciendo como respuesta “No procede” aquellas que quedaran fuera de evaluación.

Se evaluó la calidad del test original.

Se respetaron las puntuaciones de los apartados que aportan información.

Se tuvieron en cuenta los términos psicométricos y se realizó una consulta del glosario cuando fue necesario.

3. RESULTADOS

A continuación, se presentan los resultados obtenidos de la aplicación del CET-R al AULA NESPLORA. Estos resultados se presentan detallados por cada uno de los apartados e ítems que conforman el CET-R.

3.1. Resumen de la valoración del test

Tabla 1

DESCRIPCIÓN GENERAL

Características	Descripción
Nombre del test	AULA
Autor	Gema Climent Martínez y Flavio Bánterla Borzaga
Autor de la adaptación española	-
Variables(s)	Atención sostenida, atención selectiva, atención alternante, control inhibitorio, consistencia de respuesta, vigilancia, alerta, velocidad de procesamiento, tendencia a la distracción, autorregulación, canal

sensorial y actividad motora.

Áreas de aplicación Psicología clínica, Psicología educativa, Neuropsicología, Psicología forense, Psicología del tráfico, Psicología del deporte y Psicopedagogía.

Soporte

Realidad Virtual

Tabla 2

Valoración general

Características	Valoración	Puntuación
Materiales y documentación	Excelente	5
Fundamentación teórica	Excelente	5
Adaptación	Deficiente	1
Análisis de los ítems	Excelente	5
Validez: contenido	Excelente	5
Validez: estructura interna	Excelente	5
Validez: análisis del DIF	-	-
Fiabilidad: equivalencia	-	-
Fiabilidad: consistencia	Excelente	5
Fiabilidad: estabilidad	-	-
Fiabilidad: TRI	-	-
Fiabilidad: Inter-jueces	-	-
Baremos e interpretación de puntuaciones	Excelente	5

Nombre del test.

-AULA.

Nombre del test en su versión original.

-

Autor/es del test original.

Gema Climent Martínez: Licenciada en psicología, Máster en Neuropsicología, Universitat autònoma de Barcelona. Posgrado especialista en Atención temprana y Bases neuropsicológicas. Universidad de la Rioja. Directora de I+D+i en Nesplora, technology & behavior.

Flavio Bánterla Borzaga: Licenciado en ingeniería informática por la Universidad del País Vasco (UPV-EHU) Diploma en Estudios Avanzados (DEA) en inteligencia computacional. Director General en Nesplora, technology & behavior.

Autor/es de la adaptación española.

-

Editor del test en su versión original.

- Nesplora, technology & behavior.

Editor de la adaptación española:

-

Fecha de publicación del test original:

-2016

Fecha de publicación del test en su adaptación española:

-No procede.

Fecha de la última revisión del test:

-2016

Clasifique el área general de las variables que pretende medir el test:

-Neuropsicología.

Breve descripción de las variables que pretende medir el test:

-Se trata de un screening neuropsicológico sobre la red atencional que recoge puntuaciones en distintas situaciones experimentales para comparar el rendimiento y describir así el perfil atencional completo. Las variables que se pueden inferir a raíz de los resultados son:

-Atención sostenida: Es la capacidad de focalizar la atención en un punto y concentrarse a lo largo del tiempo para realizar una determinada tarea o acción.

-Atención selectiva: Es la capacidad de distinguir entre estímulos relevantes e irrelevantes para descartar información innecesaria y quedarse con lo útil.

-Atención alternante: Es la capacidad de cambiar entre distintos puntos de atención, tiene que ver con la memoria de trabajo.

-Control inhibitorio: Es la capacidad de inhibir los impulsos, la regulación del comportamiento.

-Consistencia de respuesta: Es la capacidad de ser constante en los tiempos de respuesta, la inconsistencia se asocia a fatiga, distracción o falta de regulación.

-Vigilancia y alerta: Es la capacidad de reaccionar ante un estímulo.

-Velocidad de procesamiento: Es la capacidad para responder en unos tiempos más bajos o altos.

-Tendencia a la distracción: Es la distractibilidad interna.

-Actividad motora: Es la cuantificación del movimiento motor del sujeto.

Área de aplicación:

- Psicología clínica, Psicología educativa, Neuropsicología, Psicología forense, Psicología del tráfico, Psicología del deporte y Psicopedagogía.

Formato de los ítems:

-Estímulos visuales y auditivos. Número de ítems: aciertos X, tarea 1 (180), aciertos X NO, tarea 2 (180), aciertos auditivos (174), aciertos visuales (186), aciertos totales (360)

Soporte:

-Informatizado: Realidad virtual.

Cualificación requerida para el uso del test de acuerdo con la documentación aportada:

-Categoría C: Titulados superiores en Psicología, Psiquiatría o Psicopedagogía y con experiencia profesional en diagnóstico clínico.

Descripción de las poblaciones a las que el test es aplicable:

-El test es aplicable a sujetos en la franja de edad comprendida entre 6 y 16 años, es decir, primaria y la E.S.O. Además, tiene audio para las instrucciones en los siguientes idiomas: español, disponible en los acentos (Castellano, argentino, mexicano, chileno, peruano y colombiano), euskera, catalán, inglés, disponible en los acentos (americano y británico) romaní, árabe, francés, italiano, turco y portugués. Por lo que se puede administrar a población cuya habla materna o simultánea pertenezca a esas lenguas.

Para realizar el test necesitas tener una capacidad de lenguaje, memoria, visual y auditiva que te permita comprender el mundo lo suficiente para retener las instrucciones y responder a la tarea.

Indique si existen diferentes formas del test y sus características. En el caso de que existan versiones informatizadas, describa los requisitos inusuales del hardware y software, si los hubiere, que fueran necesarios para administrar correctamente el test:

-El sistema necesario para ejecutar el test son gafas de realidad virtual: Cinemizer, Vuzix, Samsung VR u Oculus (Go, Rift, Quest 1, Quest 2) un ordenador con los siguientes requisitos mínimos: Sistema operativo: Windows 10, o Mac OS X Memoria; RAM: 2GB; Tarjeta gráfica: Intel HD Graphics o superior; Red WiFi con conexión a Internet y unos Auriculares de Diadema.

Procedimiento de corrección:

-Automatizado online.

Puntuaciones:

-Las puntuaciones directas corresponden al número de aciertos, fallos y tiempos, velocímetro y cálculos secundarios (medias o desviaciones) a estas puntuaciones.

-Estas puntuaciones se tipifican y aparece también su puntuación percentil.

Escalas utilizadas.

-Todas las puntuaciones a lo largo del informe se presentan en Puntuación Directa, Puntuación Típica y Percentil.

Posibilidad de obtener informes automatizados.

-Sí, por un lado, un informe de 12 páginas que sigue la siguiente estructura: Portada con datos sociodemográficos, fecha de administración y anotaciones en el software; Sigue con una explicación de en qué puntuaciones se explica el informe y cómo se distribuye; luego explica con gráficos todas las puntuaciones en las distintas escalas y situaciones experimentales.; tras eso, muestra una gráfica del comportamiento motor registrado; y, al final, hay una tabla resumen con todas las variables distribuidas en puntuaciones T.

Por otro lado, un informe de 14 páginas llamado "AULA school" que está diseñado como informe dinámico de carácter psicopedagógico, por lo que aparece explicado de una forma descriptiva el perfil atencional del niño en relación a sus puntuaciones y orientaciones a seguir para sacar provecho de su perfil atencional.

Éste también muestra las puntuaciones en T, Pc y Pd con gráficas, solo que en vez de ir de "rendimiento muy bajo" a "rendimiento muy alto" lo expresa en Alto, adecuado y mejorable, explicando las puntuaciones en "fortalezas" y "debilidades"

Tiempo estimado para la aplicación del test (instrucciones, ejemplos y respuestas a los ítems):

-Entre 15 y 20 minutos.

Documentación aportada por el editor:

-Manual, artículos complementarios, dossier de publicaciones, guías de interpretación, webinars, informes ejemplo y claves de producto.

Precio de un juego completo de la prueba (documentación, test, plantillas de corrección; en el caso de tests informatizados no se incluye el coste del hardware):

-Consiste en 3 planes descritos, uno es de 100€ (licencia) + 25€ (corrección única); mientras que los otros son suscripciones de hasta 1600€ (licencia anual con correcciones ilimitadas), Otras son 90€ mensuales con 7 correcciones que renuevan cada mes, 270€ trimestrales con 22 correcciones que caducan o 150€ mensuales con correcciones ilimitadas.

Todas las licencias incluyen soporte técnico y todos los test de Nesplora (AULA, AULA School, Aquarium, Ice Cream y Suite) siendo los usos de carácter global.

Todos los precios son sin IVA revisados el 06/06/2021.

-No procede.

-No procede.

-No procede.

Precio de la administración y/o corrección, y/o elaboración de informes por parte del editor:

-25€ sin IVA revisados el 06/06/2021.

Tabla 3

2.1 Aspectos generales

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Excelente	5
2.2. Calidad de la documentación aportada	Excelente	5
2.3. Fundamentación teórica	Excelente	5
2.4. Adaptación del test	Deficiente, no tiene baremos de otros países ni manuales o materiales adap-	1

tado a las lenguas que ofrece

2.5. Desarrollo de los ítems del test	Excelente	5
2.6. Calidad de las instrucciones para el participante	Excelente	5
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Excelente	5
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Excelente	5
2.10. Datos sobre el análisis de los ítems	Excelente	5

3.2. Validez**Tabla 4***2.11.1 Evidencias de validez de contenido*

Contenido	Valoración	Puntuación
2.11.1.1. Calidad de la representación del contenido o dominio	Excelente	5
2.11.1.2. Consultas a expertos	Excelente	5

2.11.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y otras variables

Tabla 5*2.11.2.1. Evidencias de validez basadas en la relación con otras variables*

Contenido	Valoración	Puntuación
2.11.2.1. Relaciones con otras variables. Diseños y/o técnicas empleados:	Correlaciones con otros tests, diferencias entre grupos, diseños experimentales.	5
2.11.2.1.2. Tamaño de las muestras	La muestra está compuesta por 1.302 sujetos con edades comprendidas entre los 5 y 22 años.	5
2.11.2.1.3. Procedimiento de selección de las muestras	Se escogieron de forma aleatoria colegios del País Vasco – Navarra	5
2.11.2.1.4. Calidad de los tests empleados como criterio o marcador	-	-
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	Excelente	5
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no relacionados	-	-
2.11.2.1.7. Resultados de la matriz multitasgomultimétodo	-	-
2.11.2.1.8. Resultados de las diferencias intergrupo	Excelente	5

Tabla 6

2.11.2.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y un criterio

Contenido	Valoración	Puntuación
-----------	------------	------------

2.11.2.2.1. Criterios empleados	-	-
2.11.2.2.2. Calidad de los criterios empleados	-	-
2.11.2.2.3. Relación temporal entre test y criterio	-	-
2.11.2.2.4. Tamaño de las muestras	-	-
2.11.2.2.5. Procedimiento de selección de las muestras	-	-
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	-	-

Tabla 7*2.11.3. Evidencias de validez basadas en la estructura interna*

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	Excelente	5
2.11.3.2. Funcionamiento diferencial de los ítems	Unidimensional	5

Tabla 8*2.11.4. Acomodaciones en la administración del test*

Contenido	Valoración	Puntuación
2.11.4. El manual del test informa sobre las acomodaciones en la administración del test	No	-

Respecto al punto 2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares (vease Tabla 5) los estudios de validez de Nesplora AULA se realizaron con pruebas de validez convergente con otros test que miden los mismos constructos. Los test con los que se realizaron las pruebas de correlación y sus respectivos promedios fueron el test Caras, con un promedio de .835, el test D-2, con un promedio de .754, el CPT de Conners (1995) con un promedio de .773 y el EDAH con un promedio situado entre .406 y .544 en la variable inatención.

A modo de ejemplo, se detallarán las puntuaciones individuales de unos de los test con los que se realizaron las pruebas de validez. En este caso el test elegido fue el CPT de Conners, y sus puntuaciones individuales fueron: Omisiones AULA/CPT Conners: 0.647, comisiones AULA/CPT Conners: 0.680, tiempo de reacción AULA/CPT Conners: 0.982, DT-Tiempo de reacción AULA/CPT Conners: 0.782.

Hay que tener en cuenta que AULA tiene 2 paradigmas evaluativos y el CPT Conners solo tiene uno. AULA presenta distractores y el Conners no. AULA se presentan los estímulos de forma visual y auditiva, en el Conners solo visual.

En el estudio de validez convergente con el EDAH se utilizaron los análisis de similitud de cosenos, obteniendo entre 0.406 y 0.544 de correlación en las variables de inatención. Estos son resultados de correlación bajos-moderados.

Por último, con respecto al análisis factorial exploratorio (Véase Tabla 7) En el análisis factorial exploratorio se obtuvo un único factor que incluía las 18 variables. Posteriormente se seleccionó de forma aleatoria al 50% de la muestra total, planteando cinco modelos posibles, desde una estructura de 5 dimensiones hasta la posibilidad de un único factor. Para la factorización se utilizó el método de ejes principales y cuando se extrajeron más de un factor se utilizó el método de rotación Direct Oblimin. Este análisis determinó que la mejor estructura sería la de un único factor con 18 variables.

Este factor explica el 33% de la varianza total y el 51% de la varianza común estimada. Se realizó también la prueba de esfericidad de Bartlett y la medida de adecuación muestral de Kaiser, Meyer y Olkin. Con estas pruebas se garantizaba que los datos fuesen adecuados para llevar a cabo el análisis factorial.

El análisis factorial confirmatorio se llevó a cabo con el 50% restante de la muestra y se analizaron los índices de ajuste al modelo, los cuales se obtuvieron mediante el método de extracción de Máxima Verosimilitud. Con respecto a los índices derivados del análisis factorial, los datos indican que existe un buen ajuste de los datos al modelo unifactorial que se planteaba, aunque el cociente X^2/gf era más elevado del que cabría esperar. En estos resultados se observó también que el valor de la raíz cuadrada media del error de aproximación era de 0.064 y el límite inferior del intervalo era de 0.059. Estos valores son aceptables para el rango en el que están, por lo que confirman que el modelo unifactorial propuesto es adecuado. Adicionalmente se comprobaron los índices del mo-

delo independiente, obteniendo siempre el modelo unifactorial mejores resultados que este.

Como conclusión, el modelo unifactorial recoge diferentes variables cognitivas que se relacionan con el diagnóstico inicial del TDAH en sus diferentes variantes, ya que con algún grupo de las 18 variables incluidas se pueden observar las alteraciones cognitivas que corresponden a cada subtipo de TDAH.

3.3. Fiabilidad

Tabla 9

2.12. Fiabilidad

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Varios coeficientes de fiabilidad para cada escala y subescala. Alfa de cronbach y ANOVA.	-

Tabla 10

2.12.2. Equivalencia formas paralelas

Contenido	Valoración	Puntuación
2.12.2.1. Tamaño de las muestras	-	-
2.12.2.2. Puesta a prueba de los supuestos de paralelismo	-	-
2.12.2.3. Promedio de coeficientes de equivalencia	-	-

Tabla 11

2.12.3. Consistencia interna

Contenido	Valoración	Puntuación
2.12.3.1. Tamaño de las muestras	Excelente	5
2.11.3.2. Coeficientes de consistencia	Coeficiente alfa de Cronbach y	5

interna presentados	Kolmogorov	
2.11.3.3. Promedio de los coeficientes de consistencia	0.97	5

Tabla 12

2.12.4. Estabilidad (test-retest)

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	-	-
2.12.4.2. Coeficientes de estabilidad	-	-

Tabla 13

2.12.5. Cuantificación de la precisión mediante TRI

Contenido	Valoración	Puntuación
2.12.5.1. Tamaño de las muestras	-	-
2.12.5.2. Coeficientes proporcionados	-	-
2.12.5.3. Tamaño de los coeficientes	-	-

Tabla 14

2.12.6. Fiabilidad inter-jueces

Contenido	Valoración	Puntuación
2.12.6.1. Tipos de coeficientes presentados	-	-
2.12.6.2. Promedio de los coeficientes	-	-

En cuanto a la fiabilidad, se utilizaron las pruebas Alfa de Cronbach y ANOVA. A continuación, se detallarán escala por escala los coeficientes alfa obtenidos y los totales de la prueba ANOVA. En todas las escalas el tamaño muestral fue de 1360 participantes.

Escala de Aciertos Auditivos, con 174 ítems, el coeficiente α de Cronbach fue de 0.963. Este coeficiente demuestra que la escala tiene una alta fiabilidad y una gran consistencia interna. En la prueba ANOVA el resultado del Análisis de la Varianza expone que la escala discrimina de forma correcta entre los niños del grupo normativo, detectando adecuadamente las diferencias interindividuales. Los ítems presentan valores aceptables en todos los casos, salvo 17 ítems que presentan valores de homogeneidad bajos.

Escala Aciertos Totales, con 360 ítems, el coeficiente α de Cronbach fue de 0.978. La escala presenta alta fiabilidad y una gran consistencia interna. El resultado de la prueba ANOVA señala que los ítems discriminan correctamente entre los niños del grupo normativo; Escala Aciertos Visuales, con 186 ítems y un coeficiente α de Cronbach de 0.966, la escala presenta una alta fiabilidad y gran consistencia interna. La prueba ANOVA indica que los ítems discriminan de forma correcta entre los niños del grupo normativo; Escala Aciertos X, con 180 ítems y un coeficiente α de Cronbach de 0.981, la escala presenta una alta fiabilidad y una gran consistencia interna. La prueba ANOVA indica que los ítems discriminan de forma correcta entre los niños del grupo normativo. En esta escala los ítems 124 y 136 presentan un índice de homogeneidad más bajo del esperado; Escala Aciertos XNO, con 180 ítems y un coeficiente α de Cronbach de 0.972, la escala presenta una alta fiabilidad y una excelente consistencia interna. La prueba ANOVA indica que los ítems discriminan de forma correcta entre los niños del grupo normativo.

3.4. Baremación e interpretación de las puntuaciones

Tabla 15

2.13.1. Interpretación normativa de las puntuaciones

Contenido	Valoración	Puntuación
2.13.1.1. Calidad de las normas	Excelente	5
2.13.1.2. Tamaño de las muestras	Excelente	5
2.13.1.3. Aplicación de estrategia	Sí	5

de tipificación continua

2.12.1.4. Procedimiento de selección de las muestras	Aleatorio, con muestras representativas de la población objetivo	5
2.12.1.5. Actualización de baremos	Excelente	5

Tabla 16*2.13.2. Interpretación referida a criterio*

Contenido	Valoración	Puntuación
2.13.2.1. Adecuación del establecimiento de los puntos de corte	-	-
2.13.2.2. Si se utiliza el juicio de expertos para establecer los puntos de corte, indique el procedimiento empleado para fijar el estándar	-	-
2.13.2.3. Si se utiliza el juicio de expertos para establecer los puntos de corte, indique cómo se ha obtenido el acuerdo interjueces	-	-
2.13.2.4. Indique el valor del coeficiente de acuerdo interjueces (e.g., Kappa o ICC)	-	-

El resultado de las diversas pruebas que se realizaron para la baremación del test concluyeron en 14 grupos de baremos, por sexos y edades. Estos resultados se baremaron en función de los destinatarios y los usuarios del test, quedando baremados en escalas de percentiles y en Escalas-T. Así mismo se normalizaron las escalas T para que su interpretación fuese más sencilla.

En el manual especifican que los baremos de puntuaciones normalizadas deben de evitarse debido a que ninguna de las variables sigue una distribución normal salvo una. Esto se comprobó mediante la prueba Z de Kolmogorov-Smirnov.

El procedimiento que se siguió fue el siguiente:

Se realizó una prueba de homocedasticidad para el estudio de la diferencia por sexos y edades. Para el estudio de las diferencias por sexo se utilizó la prueba F, resultando ésta en que 22 de las 54 variables no cumplían el supuesto de homocedasticidad. Para el estudio de las diferencias por edad se utilizó la prueba de Levene, siendo el resultado de esta que solo dos variables cumplían con el supuesto de homocedasticidad. Antes de comenzar la baremación se debían de estudiar las diferencias interindividuales, ya que los resultados de estas pruebas determinarán la cantidad de baremos necesarios.

Para el estudio de las diferencias en función de la edad y el sexo del grupo normativo se llevaron a cabo diversos análisis paramétricos y no paramétricos, ya que se violan los supuestos de homocedasticidad y el de normalidad. Se realizó una prueba “t” y una prueba “U” de Mann-Whitney para las diferencias por sexos y una ANOVA de un factor y la prueba de Kruskal-Wallis para las diferencias por edad. Además, se sumó la prueba “W” de Wilcoxon a la “U” de Mann-Whitney. Los resultados de estas dos pruebas fueron coincidentes, aunque los resultados de la prueba “U” y “W” son más adecuados en función de las características de los datos.

Respecto a la edad, se encontraron diferencias en función de la edad en todas las variables salvo en 7, por lo que se realizó el mismo estudio, pero esta vez con una prueba no paramétrica (Kruskal Wallis), con la que encontramos diferencias en todas las variables en función de la edad.

Una vez obtenidos los resultados de la aplicación del test al grupo normativo se ejecutaron la prueba de Kolmogorov-Smirnov para comprobar la normalidad de las distribuciones de las puntuaciones en las distintas escalas y una prueba de F para contrastar la igualdad de varianzas. Los resultados fueron obtenidos a un nivel de confianza del 95%, y estos declararon que en la mayoría de los casos no se cumplía el supuesto de homocedasticidad. Además, tampoco se podían llevar a cabo las pruebas paramétricas ya que no se cumplía el supuesto de normalidad en la distribución. Aun así, se llevaron a cabo tanto pruebas paramétricas como no paramétricas. Las pruebas paramétricas demostraron tener la robustez suficiente ante la violación de los supuestos de homocedasticidad y distribución normal, ya que los resultados obtenidos siempre fueron coincidentes en ambos tipos de pruebas.

En cuanto al estudio de las diferencias de medias entre los diferentes grupos, las pruebas realizadas fueron: Pruebas “t” y “U” de Mann-Whitney para el estudio de diferencias entre sexos, ANOVA, MANOVA y prueba de Kruskal-Wallis para el estudio de las diferencias por edades y sexos.

Estas pruebas encontraron diferencias estadísticamente significativas entre hombres y mujeres en todos los grupos de edad y en casi todas las variables baremadas. Además, también se encontraron diferencias estadísticamente significativas entre las edades de 6,7,8,9,10,11 y mayores de 12 años. De 12 años en adelante no se encontraron diferencias.

3.5. Valoración final

3.1. Con una extensión máxima de 1000 palabras, exprese su valoración del test, resaltando sus puntos fuertes y débiles, así como las recomendaciones acerca de su uso en diversas áreas profesionales. Indique asimismo cuáles son las características de la prueba que podrían ser mejoradas, carencias de información en la documentación, etc.

AULA Nesplora presenta una excelente justificación estadística, sus puntos fuertes son la increíble consistencia interna que tiene, que emite informes automatizados y gracias a su entorno presenta una gran validez ecológica que elimina el sesgo de evaluador y ambiental. Con respecto a puntos débiles, dependes de un hardware externo que es caro, no está adaptado a personas con discapacidad ni presenta todo el material disponible en las distintas lenguas que ofrece.

También logra medir los mismos valores que otros tests informatizados, pero en más situaciones experimentales distintas, haciendo que sus resultados sean únicos y más completos. Por último, gracias a la tecnología consiguen añadir una medida de actividad motora, algo totalmente revolucionario.

Gracias al informe de AULA se pueden apoyar diagnósticos psiquiátricos, neurológicos y psicológicos en el ámbito clínico, teniendo además puntuaciones muy fiables sobre las que fijarse para establecer terapias y pudiendo usarse el mismo test para monitorizar la terapia y realizar el seguimiento.

Por último, el informe de AULA School presenta un enfoque muy adecuado para mejorar la comunicación con el entorno familiar y educativo, estas distintas posibilidades de informes considerando distintas disciplinas resultan muy apropiadas para aprovechar al máximo la herramienta.

4. DISCUSIÓN

Tal y como hemos comentado en el manual de NESPLORA se aportan evidencias de fiabilidad y de validez interna y externa, pero además se ha realizado una revisión de la literatura para obtener otras evidencias de las propiedades psicométricas de esta herramienta de evaluación, y completar la información disponible sobre este test en términos de evidencias de validez convergente, sensibilidad y otros usos de NESPLORA.

4.1. Validez convergente

En este apartado se describen estudios experimentales publicados en revistas científicas por diversos investigadores. Las evidencias de validez convergente se obtienen al evaluar en una misma muestra de personas la variable objeto de medida con diferentes herramientas, siendo una el test diseñado para medirla, del que se pretende obtener evidencias de validez, y analizar el grado de relación entre esas medidas.

AULA es analizada junto a diversos CPT como CONNERS, Díaz-Orueta et al., (2014) evaluaron 57 sujetos para analizar ambas pruebas. Los resultados muestran que hay correlación significativa en estímulos visuales, aunque no tiene estímulos auditivos, muestra también correlación con estos, con el paradigma no-X tanto en auditivos como en visuales, y en el paradigma X o GO (Más bajo aún en comisiones en este último) siendo la correlación más alta cuando no hay distractores en AULA. Hay que tener en cuenta que pueden compararse pocas variables, ya que el CPT Connors solo contiene estímulos visuales, no tiene distractores, usa solo un paradigma (No-X o No-Go, AULA usa dos -X o Go- en dos tareas diferenciadas) y el tiempo de presentación de estímulos de Connors es fijo mientras que el de AULA se adapta al tiempo de respuesta del sujeto.

En otro estudio (Zulueta et al., 2013) demuestra que AULA indica correlaciones mayores de 0.6 entre los resultados obtenidos en AULA y el test de CARAS en variables de atención selectiva, atención sostenida e impulsividad en 62 niños.

4.2. Sensibilidad

Tras la publicación del estudio normativo del test AULA (Iriarte et al., 2012), los autores comentaron en las conclusiones que futuros estudios arrojarían datos objetivos sobre la sensibilidad de la prueba, los más relevantes, siguen comparando la prueba con otros test demostrando también su validez convergente, como en éste estudio de Rodríguez et al., (2018) incluido en un metaanálisis de (Parsons et al., 2019) en el que comparan otro test informatizado tipo CPT, el test TOVA (Leark et al., 2020), y AULA con una muestra de 338 niños (31.95% inatentos, 15.38% impulsivos hiperactivos, 22.78% combinados y 29.88% sin dificultades atencionales) para observar la fiabilidad de ambos, demostrando que AULA distingue mejor los subtipos de TDAH y además diferenciar mejor entre estudiantes con y sin TDAH que el test TOVA.

Otros, más centrados en el diagnóstico que en la convergencia de otras herramientas como el de Areces, et al., (2018b) en el que se analizó a 117 sujetos, 90 varones y 27 mujeres todos diagnosticados con TDAH para verificar si Nesplora AULA distinguía entre el subtipo impulsivo hiperactivo, intento y combinado. Los resultados muestran que cada variable del test permitía hacer un diagnóstico diferencial adecuado para cada subtipo en distintas condiciones contextuales.

En otro estudio (Fernández Fernández, et al., 2012), en la valoración del TDAH en 40 pacientes de rango de edad entre 6-16 siendo el 65% varones. Se muestra que el 100% pudo realizar el estudio sin alteraciones, el 97% presentó resultados compatibles con la existencia de excesivos niveles de inatención, inquietud motora o impulsividad. En el 3% restante pudo confirmarse la existencia de altas capacidades intelectuales que supusieron un sesgo en la realización del estudio. Los resultados correlacionaron de forma directamente proporcional con los resultados de las escalas de valoración clínica. Además, los autores añadieron que el test gozaba de fácil cumplimentación, alto valor predictivo positivo y fiabilidad para diagnóstico del TDAH, aunque existen algunos grupos de población en los que los resultados se ven afectados por el nivel intelectual del individuo.

En la misma línea Rufo et al., (2012), buscan determinar la capacidad de discriminación de la herramienta con niños diagnosticados de TDAH, 62 sujetos, frente a un grupo control de otros 62 sujetos. Para ello utilizan todas las variables y obtienen una clasifi-

cación correcta del 93.5%, concluyendo que tanto la sensibilidad como su poder diagnóstico resultan excelentes.

Por último, con una muestra compuesta por 407 niños, 272 niños y 165 niñas de entre 6 y 16 años, estando 213 diagnosticados con TDAH, de entre los cuales 105 son tipo desatentos y 108 tipo combinado, siendo los 194 restantes controles Zulueta et al., (2018) aportan evidencias de adecuada validez externa del instrumento AULA que permite realizar clasificaciones correctas de sujetos sanos y patológicos confirmándose que proporciona información adicional para el diagnóstico de TDAH, además comentan que esto es esencial para el posterior diseño de intervenciones clínicas eficaces.

AULA no solo nos permite dar diagnósticos más objetivos y sensibles respecto a nuestras sospechas, sino diferenciales, con otra muestra de 88 estudiantes de los cuales 50 están diagnosticados con TDAH, siendo 66 son varones, Areces, et al., (2018a) muestran que los estudiantes que tienen baja memoria de trabajo y velocidad de procesamiento en WISC también tienen una peor performance en AULA. El estudio sugiere que los problemas atencionales pueden ser la fuente del bajo rendimiento en tareas que incluyen memoria de trabajo y, los tiempos de respuesta bajos con tareas que tienen el tiempo como parte del rendimiento. Esto indica la importancia de obtener resultados de distintas variables cuando consideramos problemas cognitivos de cara a perfilar mejor los perfiles de los pacientes.

4.3. Usos fuera del ámbito de la psicología clínica

Si bien la evaluación psicológica suele ser objeto del ámbito psicológico, también hay otras disciplinas que pueden beneficiarse de sus resultados para mejorar la eficacia de sus tratamientos. En el caso de la psiquiatría, encontramos en la literatura cómo se utilizan las evaluaciones para probar la eficacia de distintos tratamientos farmacológicos en personas evaluadas.

Según García y González (2015), es fundamental el manejo del TDAH conjuntamente con la psicoeducación, la intervención psicopedagógica y psiquiátrica, utilizando farmacoterapia se demuestra que aumentando la tasa de respuesta y mejoran muchos de los síntomas. Las escalas también permiten valorar la eficacia de los medicamentos empleados. de igual forma posibilitan el seguimiento de la evolución del paciente que está bajo tratamiento (Herrán et al., 2014).

En otro estudio (Díaz-Orueta et al., 2016), buscaron analizar la eficacia de la lisdexanfetamina en la mejora sintomática conductual y cognitiva del trastorno por déficit de atención/hiperactividad en una muestra de 85 pacientes de 6-16 años con diagnóstico clínico de TDAH y que asistían a tratamiento en consulta de neuropsiquiatría. En el estudio encontraron mejorías muy significativas en la atención selectiva y sostenida, calidad del foco atencional, hiperactividad, mejoría moderadas en la impulsividad, y una incidencia casi nula en la velocidad de procesamiento.

5. CONCLUSIONES

Considero, para finalizar, que los resultados de las aplicaciones del CET-R permiten al usuario contar con información detallada acerca de la calidad psicométrica de los instrumentos de evaluación psicológica aportando una mayor garantía a la hora de seleccionar y utilizar una determinada prueba frente a otras (Hidalgo et al., 2019). Además, AULA Nesplora resulta tener una puntuación excelente, siendo un test recomendable a la hora de realizar evaluaciones psicológicas de la atención.

Con respecto a los CPT en realidad virtual, la necesidad de hacer controles de calidad aumenta de la misma manera que los test van siendo publicados, en este artículo (García et al., 2014) se muestra la necesidad de entornos ecológicos donde realizar medidas fiables que equiparen el comportamiento del niño a un contexto real, por lo que sería interesante proponer para futuras revisiones CET-R los otros test de NESPLORA.

5.1. Limitaciones

Una limitación encontrada en este trabajo fue la búsqueda bibliográfica realizada. Al ser un estudiante de Grado el acceso a las bases de datos bibliográficas está bastante limitado, ya que muchos de los trabajos de investigación identificados en la búsqueda no se pudieron recuperar debido a que estaban restringidos o eran de pago.

Otra limitación encontrada fue la novedad del test AULA, más bien la novedad de los test informatizados. Estos al ser tan actuales aún no tienen la trayectoria de trabajo que tienen los test clásicos ni las revisiones de estos, por lo que poder compararlos y evaluarlos con el CET-R ha sido algo más complejo. Por otro lado, algunos aspectos a mejorar del CET-R mencionado en revisiones anteriores han sido también parte de los que he encontrado durante la aplicación. Un aspecto en el que coincido con Hidalgo et al., (2019) en el caso de baremos, el modelo da una puntuación más alta cuando se utiliza un amplio

rango de baremos en vez de un único baremo adecuado para la población objetivo. Además, ya en esta revisión mencionada comentan que algunas editoriales opinan que esto podría llevar a introducir artificialmente baremos adicionales que pueden ser innecesarios con el fin de obtener una mejor puntuación. También se señala que algunos de los criterios son excesivamente exigentes en aplicaciones realistas. Finalmente, la subjetividad inherente a las valoraciones que se dan, especialmente las cualitativas también se valora como un aspecto negativo en el proceso. Es decir, se debiera introducir una justificación más exigente o con criterios más claros para estas valoraciones.

Otros aspectos a mejorar que también he observado se comentan en la evaluación coordinada por Fonseca-Pedrero y Muñiz, (2017) en la que comentan algunos de los ítems del CET-R parecen ser difíciles de valorar por los expertos en anteriores revisiones, por tanto, demandan incorporar unas pautas generales para la utilización del CET-R, así como una mayor explicación de algunos de los ítems. No obstante, parece ser que esta dificultad continúa otras evaluaciones. La dificultad a la hora de interpretar, analizar y puntuar determinados ítems como los referidos al análisis de ítems o la validez criterial, y, en ambos casos, se encontraron diferencias sustanciales entre las puntuaciones y comentarios enviados al coordinador por los revisores.

En el futuro, es probable que se necesite otra revisión del CET-R para incluir ciertos aspectos que no evalúa sobre las herramientas como éstas, como su validez ecológica, que, cada vez, serán más habituales en nuestro catálogo, incluso (Fonseca-Pedrero y Muñiz, 2017) mencionan en que el CET-R pudiera tener una versión web que pueda ser utilizada por los distintos profesionales. No obstante, en el actual CET-R ya se hace mención al formato digital de test, pero son muchos los ítems y parámetros que no se adaptan al formato clásico de test y que no hemos podido resolver. En un futuro seguramente este tipo de test esté mucho más implementado en la evaluación y el CET-R volverá a tener una revisión que incluya aspectos de este tipo de test que en la actualidad están fuera de la batería de ítems.

6. REFERENCIAS

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>

- Areces, D., Dockrell, J., García, T., González-Castro, P. y Rodríguez, C. (2018a). Analysis of cognitive and attentional profiles in children with and without ADHD using an innovative virtual reality tool. *PLOS ONE*, 13(8). <https://doi.org/10.1371/journal.pone.0201039>
- Areces, D., Rodríguez, C., García, T., Cueli, M. y González-Castro, P. (2018b). Efficacy of a Continuous Performance Test Based on Virtual Reality in the Diagnosis of ADHD and Its Clinical Presentations. *Journal of Attention Disorders*, 22(11), 1081-1091. <https://doi.org/10.1177/1087054716629711>
- Climent, G. y Banterla, F. (2016). *Nesplora AULA Manual*. https://repository.nesplora.com/docs/manual_nesplora_aula.pdf
- Conners, C. (1995) Conner's Continuous Performance Test, computer program (CPT). Toronto: Multi-health Systems.
- Delgado, G. y Moreno, I. (2012). Virtual Reality Applications in Attention Deficit Disorder with Hyperactivity: An Approximation. *Annuary of Clinical and Health Psychology* 8, 29-37.
- Díaz Orueta, U., Fernández Fernández, M. A., Morillo, M. D. y Climent, G. (2016). Eficacia de la lisdexanfetamina en la mejora sintomática conductual y cognitiva del trastorno por déficit de atención/ hiperactividad: Tratamiento monitorizado mediante el test AULA Nesplora de realidad virtual. *Revista de Neurología*, 63(1), 19-27. <https://doi.org/10.33588/rn.6301.2015488>
- Díaz-Orueta, U., Garcia-López, C., Crespo-Eguílaz, N., Sánchez-Carpintero, R., Climent, G. y Narbona, J. (2014). AULA virtual reality test as an attention measure: Convergent validity with Conners' Continuous Performance Test. *Child Neuropsychology*, 20(3), 328-342. <https://doi.org/10.1080/09297049.2013.792332>
- Dolores Hidalgo, M. y Hernández, A. (2019). SEXTA EVALUACIÓN DE TESTS EDITADOS EN ESPAÑA: RESULTADOS E IMPACTO DEL MODELO EN DOCENTES Y EDITORIALES. *Papeles del Psicólogo - Psychologist Papers*, 40(1), 21-30. <https://doi.org/10.23923/pap.psicol2019.2886>
- Fonseca-Pedrero, E. y Muñiz, J. (2017). Quinta evaluación de Tests editados en España: Mirando hacia atrás, construyendo el futuro. *Papeles del Psicólogo - Psychologist Papers*, 38(3), 161-168. <https://doi.org/10.23923/pap.psicol2017.2844>
- García, J. M. y González, J. J. (2015). Guía de Algoritmos en Pediatría de Atención Primaria. Trastorno por Déficit de Atención con Hiperactividad (TDAH). *AEPap*, algoritmos.aepap.org
- García, T., González-Castro, P., Areces, D., Cueli, M. y Rodríguez, C. (2014). Funciones ejecutivas en niños y adolescentes: implicaciones del tipo de medidas de evaluación empleadas para su validez en contextos clínicos y educativos. *Papeles del Psicólogo*, 35(3), 215-223.
- Hernández, A., Ponsoda, V., Muñiz, J., Prieto, G. y Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 37, 192-197.

- Iriarte, Y., Díaz-Orueta, U., Cueto, E., Irazustabarrena, P., Banterla, F. y Climent, G. (2012). AULA—Advanced Virtual Reality Tool for the Assessment of Attention: Normative Study in Spain. *Journal of Attention Disorders*, 20(6), 1-27. <https://doi.org/10.1177/1087054712465335>
- Kessels, R. P. C. (2018). Improving precision in neuropsychological assessment: Bridging the gap between classic paper-and-pencil tests and paradigms from cognitive neuroscience. *The Clinical Neuropsychologist*, <https://doi.org/10.1080/13854046.2018.1518489>
- Leark, R., Dupuy, T., Greenberg, L., Kindschi, C. y Hughes, S. (2020). T.O.V.A.® Professional Manual. <http://files.tovatest.com/documentation/8/Professional%20Manual.pdf>
- Muñiz, J. y Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. <https://doi.org/10.7334/psicothema2018.291>
- Olea, J., Abad, F. J. y Barrada, J. R. (2010). Tests informatizados y otros nuevos tipos de tests. *Papeles del Psicólogo*, 31(1), 94-107. <http://www.cop.es/papeles>
- OMS. (1992). *CIE-10. Décima revisión de la clasificación internacional de las enfermedades. Trastornos mentales y del comportamiento. Descripción clínicas y pautas para el diagnóstico*. Méditor.
- Parsons, T. D., Duffield, T. y Asbee, J. (2019). A Comparison of Virtual Reality Classroom Continuous Performance Tests to Traditional Continuous Performance Tests in Delineating ADHD: A Meta-Analysis. *Neuropsychology Review*, 29(3), 338-356. <https://doi.org/10.1007/s11065-019-09407-6>
- Rodríguez, C., Areces, D., García, T., Cueli, M. y González-Castro, P. (2018). Comparison between two continuous performance tests for identifying ADHD: Traditional vs. virtual reality. *International Journal of Clinical and Health Psychology*, 18(3), 254-263. <https://doi.org/10.1016/j.ijchp.2018.06.003>
- Rodríguez-Barranco, M., Gil, F., Hernández, A. F., Alguacil, J., Lorca, A., Mendoza, R., Gómez, I., Molina-Villalba, I., González-Alzaga, B., Aguilar-Garduño, C., Rohlman, D. S. y Lacasaña, M. (2016). Postnatal arsenic exposure and attention impairment in school children. *Cortex*, 74, 370-382. <https://doi.org/10.1016/j.cortex.2014.12.018>
- Viladrich, C., Doval, E., Penelo, E., Aliaga, J., Espelt, A., García-Rueda, R. y Angulo-Brunet, A. (2021). Octava evaluación de test editados en España: una experiencia participativa. *Papeles del Psicólogo - Psychologist Papers*, 42(1), 1-9. <https://doi.org/10.23923/pap.psicol2020.2937>
- Zulueta, A., Díaz-Orueta, U., Crespo-Eguilaz, N. y Torrano, F. (2018). Virtual Reality-based Assessment and Rating Scales in ADHD Diagnosis. *Psicología Educativa*, 25(1), 13-22. <https://doi.org/10.5093/psed2018a18>
- Zulueta, A., Iriarte, Y., Díaz-Orueta, U., Sebastián, S. y Climent, G. (2013). AULA Nesplora: avance en la evaluación de los procesos atencionales. Estudio de la validez convergente con el test de percepción de diferencias “caras” (versión ampliada). *ISEP Science*, 4, 3-10.