

Diseño y validación de una prueba de selección para controladores de tráfico marítimo basada en la medida de la conciencia situacional



José R. Cordón^{a*}, Pedro Ramiro Olivier^b, Manuel A. García Sedeño^b, and Jorge Walliser Martín^a

^aSociedad Estatal de Salvamento y Seguridad Marítima-Universidad de Cádiz, Spain

^bUniversidad de Cádiz, Spain

INFORMACIÓN ARTÍCULO

Manuscrito recibido: 24/07/2013

Revisión recibida: 04/02/2014

Aceptado: 24/06/2014

Palabras clave:

Test de Conciencia Situacional

Controlador

Piloto

Servicio de tráfico marítimo

Validación

RESUMEN

En este artículo se resalta la necesidad de contar con un test fiable para la selección de futuros operadores en los servicios de tráfico marítimo. Se presenta el proceso de creación del Situation Awareness Test for Vessel Traffic Services (SAT-VTS) [Prueba de Conciencia Situacional para Servicios de Tráfico Marítimo]. Se expone la base teórica sobre la que descansa la prueba, que es la conciencia situacional, y el proceso de su construcción y validación, para lo que se utilizó una prueba de jueces expertos y posteriormente una muestra de posibles futuros candidatos ($N = 138$), alumnos de Náutica de la Universidad de Cádiz. Se detallan los niveles de fiabilidad de la prueba ($\alpha = .864$) así como otras características psicométricas.

© 2014 Colegio Oficial de Psicólogos de Madrid. Producido por Elsevier España, S.L. Todos los derechos reservados.

Design and validation of a screening test for vessel traffic services operators based on situational awareness assessment

ABSTRACT

This article highlights the need of a reliable test for the selection of future operators in vessel traffic services. It describes the theoretical basis of the test that lies in situational awareness and the process of building and validation of the Situation Awareness Test for Vessel Traffic Services (SAT-VTS) by means of a test of expert judges, and then a sample of possible future candidates ($N = 138$), students of the Sea Navigation School at Cadiz University (Spain). The reliability of the test ($\alpha = .864$) and other psychometric characteristics are detailed.

© 2014 Colegio Oficial de Psicólogos de Madrid. Production by Elsevier España, S.L. All rights reserved.

La conciencia situacional y su aplicación al control de tráfico marítimo

¿Qué hace que una persona rinda bien en un determinado contexto? Esta es una de las preguntas clave que se plantean en el entorno de la psicología de los recursos humanos. En general se podría responder que el rendimiento de la persona queda determinado por diversos factores relacionados con la personalidad del sujeto, la tarea a desempe-

ñar, las circunstancias particulares en que el individuo se desenvuelve, etc. Esta misma cuestión, particularizada para el contexto marítimo, podría expresarse como ¿qué hace de un marino un buen marino, o un marino excelente? En el contexto de los servicios de tráfico marítimo se expresaría como ¿qué hace que una persona rinda adecuadamente en el control de tráfico marítimo? Hasta el momento no puede decirse que exista una respuesta concreta a estas cuestiones.

En 1993 se desarrollan en España los primeros servicios de tráfico marítimo gestionados por la entidad público empresarial SASEMAR (Sociedad de Salvamento y Seguridad Marítima), empresa pública adscrita al Ministerio de Fomento.

Desde un primer momento, en los procesos de selección y evaluación de los aspirantes a controladores u operadores de servicios de

*La correspondencia sobre este artículo debe enviarse al primer autor al Departamento de Psicología, Facultad de Ciencias de la Educación, Campus Río San Pedro, 11519 Puerto Real, Cádiz, España. E-mail: joserc@sasemar.es

tráfico marítimo dicho organismo ha utilizado pruebas estándar de atención y personalidad. Dicho procedimiento de evaluación ha sido insatisfactorio, toda vez que ha producido resultados indeseados tales como la admisión de candidatos que luego han resultado inadecuados para el puesto y, naturalmente, el resultado opuesto como es la eliminación de candidatos válidos para el desempeño de las tareas.

Como consecuencia de las disfunciones detectadas, un grupo de profesionales de SASEMAR, con más de 15 años de experiencia en el campo de los Servicios de Tráfico Marítimo y al menos 10 años de experiencia en el ámbito de la navegación marítima y vinculados a la Universidad de Cádiz, promueven con el apoyo de la Dirección de Operaciones de dicho organismo una serie de estudios que permitan definir un nuevo modelo de selección de personal que ofrezca resultados acordes a las necesidades específicas del servicio que ofrece a la comunidad marítima. Estos servicios se verán incrementados en un futuro por la obligación contemplada en ley (BOE, 2011) de que todos los puertos del Estado provean de un servicio de control de tráfico marítimo en sus zonas de responsabilidad; es decir, nace de una necesidad real actual y con perspectiva de crecimiento futuro y un aumento del nivel de compromiso con el incremento de los estándares de seguridad en la navegación marítima en el ámbito de los espacios marítimos próximos a la costa.

La conciencia situacional: definición y concepto

La conciencia situacional [Situation Awareness] (SA en adelante) es el proceso de creación y comprensión de un mapa o representación mental del entorno del sujeto, especialmente en sistemas complejos y con sobrecarga estimular, a través de la selección de la información recibida, su posterior elaboración y, finalmente, la toma de decisiones pertinentes.

Se entiende como conciencia situacional la percepción de elementos en el entorno, dentro de una ventana espacio-temporal definida, la comprensión de su significado y la proyección de su status en un futuro cercano (Endsley, Farley, Jones, Midkiff y Hansman, 1998).

Si bien en un principio el paradigma del procesamiento de la información ha dominado el trabajo teórico sobre este constructo, con el crecimiento del paradigma naturalista el interés se ha ido desplazando hacia los procesos de toma de decisiones. Todo ello ha coincidido con un importante desarrollo de la investigación en torno al concepto (McNeese, 1999). En cualquier caso, consideramos que ambas aproximaciones no tienen por qué ser excluyentes, pudiendo el proceso de toma de decisiones estar englobado en el procesamiento de la información.

La teoría de tres niveles de Endsley

El desarrollo del presente trabajo toma como base teórica la teoría de tres niveles de la conciencia situacional (Endsley, 1995), por su sencillez, facilidad de adaptación al campo de la Marina Civil, amplia difusión en otros entornos y sistemas y la abundante literatura existente al respecto. En esencia esta teoría es aplicable a cualquier tarea que requiera el seguimiento de una serie de eventos.

El modelo de Endsley se estructura en tres niveles jerárquicos de conciencia situacional, cada uno de los cuales es necesario, pero no suficiente, precursor del nivel siguiente y superior. El modelo se basa en el establecimiento de una cadena de procesamiento de la información que, dividida en tres niveles, comienza con la percepción de la misma, prosigue con la interpretación y finaliza con la predicción.

Partiendo del nivel inferior, el esquema del proceso sería como sigue:

Nivel 1 SA: percepción de los elementos del entorno. Es el nivel inferior y Endsley lo relaciona con la percepción de la información recibida a través de la instrumentación disponible, del comportamiento de la nave, de la información asociada a otros sujetos del entorno, a otras naves en las proximidades, al terreno y al control de

tráfico. En este estadio no se interpretan los datos, sólo se almacenan en forma "bruta", sin elaborar. Algún dato puede servir para confirmar el estado de una variable, pero sin ser integrado todavía por el individuo.

Nivel 2 SA: comprensión de la situación actual. Se alcanza a través de la percepción de los elementos del entorno. En este estadio no es necesario que se produzca una comprensión relevante para el desarrollo de su tarea por parte del piloto u operador. Es el nivel en el que se produce el proceso de integración de los datos. Según establece Endsley, la conciencia situacional va más allá de la percepción, contemplando también el proceso mediante el cual los sujetos combinan, interpretan, almacenan y retienen información. De esta manera la conciencia situacional incluye no solo la percepción de la información del entorno, sino también la integración de múltiples piezas de información y la determinación de su relevancia en función de los objetivos del sujeto.

Nivel 3 SA: predicción del estatus futuro. Este es el nivel más elevado de conciencia situacional, que se asocia con la habilidad para proyectar el estatus futuro de los elementos del entorno. La precisión en la predicción depende enormemente de la información obtenida en los niveles 1 y 2. La anticipación proporciona al sujeto el tiempo suficiente para resolver conflictos y planificar su actuación para alcanzar sus objetivos. Los estudios de campo realizados por Endsley, Sollenberger y Stein (2000) establecen que la seguridad en las predicciones que hace el sujeto está en gran medida determinada por su experiencia, hasta el punto de considerar que la confianza en las predicciones realizadas "es señal de que se trata de un operador experto" (p. 6).

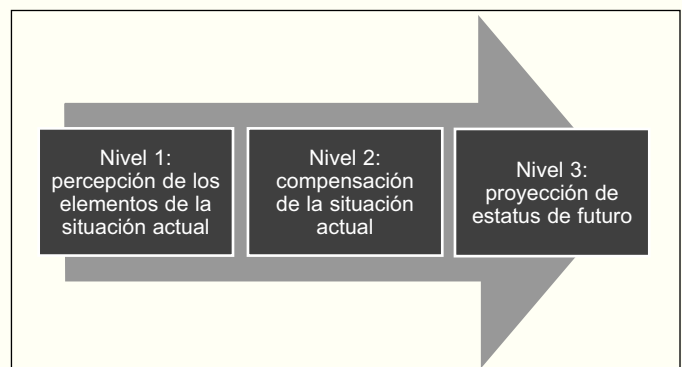


Figura 1. Modelo de los tres niveles de conciencia de la situación de Endsley (1995).

Según señala Endsley, la SA está influenciada tanto por factores asociados a la tarea como por otros asociados a las características específicas del individuo. Así se explica de qué modo diferentes personas en contextos similares llegan a diferentes conclusiones según su habilidad, entrenamiento y experiencia.

SA en otros ámbitos de la psicología

El mayor desarrollo y aprovechamiento del concepto de conciencia situacional se ha dado en el campo del control de tráfico aéreo y en aviación en general. Con el aumento del tráfico aéreo se vio que las exigencias de los sistemas de control sobrepasaban ampliamente la capacidad humana de mantenerse atento a la situación (Jensen, 2013). Se observa un importante paralelismo entre esta situación y la del campo del control del tráfico marítimo. En este se observa además la proliferación de nuevas interfaces que muestran solo información relevante para el entorno de navegación y en muchos casos hace que disminuya la atención que el oficial de guardia de navegación debe prestar a la respuesta del buque o a la situación real que le rodea.

En cualquier caso, la importancia de mantener un elevado nivel de conciencia de la situación no queda limitada al campo de la avia-

ción. Autores como Kaber y Endsley (1998) propugnan la extensión de este concepto a otros campos más allá de la aviación. De hecho su aplicación es de capital importancia en ámbitos como los sistemas de control industrial, el campo médico, el tráfico rodado o el marítimo.

Entre los autores destaca Endsley tanto por el volumen de su producción científica como por la influencia de ésta en la literatura sobre la conciencia situacional. Aunque sus investigaciones se centran sobre todo en la aviación civil y militar, sin embargo la base teórica por ella desarrollada es sin duda la más utilizada en investigaciones realizadas en otros ámbitos de interés.

Otros autores han llevado a cabo importantes revisiones de la literatura sobre la conciencia situacional (p. ej., Banbury, Tremblay, Rousseau, Forbes y Breton, 2008). En este sentido cabe destacar el número especial que la revista *Safety Science* dedica a la investigación que ha generado este constructo (Salmon y Stanton, 2013). Sin pretender hacer una revisión exhaustiva del concepto, que no es el propósito del presente artículo, debe señalarse que existen referencias de la adaptación del constructo de conciencia situacional a campos tan dispares como:

- Combate de infantería (Banbury et al., 2008; Endsley, Bolstad, Jones y Riley, 2003; Jones, Bolstad, Riley, Endsley y Shattuck, 2003).
- Combate aéreo (Dennehy, 1997; Endsley, 1997).
- Control de tráfico aéreo (Ahlstrom, 2007; Baumgartner, Gottesheim, Mitsch, Retschitzegger y Schwinger, 2010; Denford, Steele, Roy y Kalantzis, 2004; Farley, Hansman, Endsley, Amonlirdviman y Vigeant-langlois, 1998; Wickens, 2002).
- Aviación civil (Bolstad, Endsley, Costello y Howell, 2010; Denford et al., 2004; Uhlarik y Comerford, 2002).
- Mantenimiento e industria (Bernard, Ims, Carver, Cauvin y Nolet, 2002; Endsley y Robertson, 2000; Itoh y Inagaki, 1996).
- Energía nuclear (Hallbert, 1997; Hogg et al., 1995; Licao, Pengcheng y Li, 2010).
- Automovilismo (Salmon, Young y Cornelissen, 2013; Schömig y Metz, 2013; Underwood, Ngai y Underwood, 2012; Walker, Stanton y Chowdhury, 2013; Young, Salmon y Cornelissen, 2013).
- Enfermería (Stubblings, Chaboyer y McMurray, 2012).
- Emergencias y medicina de emergencias (Blandford y Wong, 2004; Javed, Norris y Johnston, 2011).
- Anestesiología (Zhang et al., 2002).
- Para finalizar, el campo donde menos ha sido aplicado el concepto, y que es el que nos ocupa, la Marina Civil. Centrada en los servicios de control de tráfico marítimo (VTS) tenemos la tesis doctoral de J. W. F. Wiersma (2010) en la Universidad de Delft, Holanda, donde se desarrolla un sistema para medir el rendimiento de los operadores de VTS para la autoridad portuaria de Rotterdam.

A la vista de las anteriores referencias se puede observar cómo se ha ido extendiendo la aplicación del constructo a diversos entornos, todos ellos de gran complejidad y con importantes implicaciones en los procesos de toma de decisión.

Aplicación de la conciencia situacional a la construcción de la prueba

El ámbito de los servicios de tráfico marítimo (en adelante VTS, siglas en inglés de Vessel Traffic Service) presenta un entorno laboral de gran sobrecarga estimular. La tarea a desarrollar está caracterizada por la intervención de un gran número de factores que condicionan el análisis que el sujeto, el operador de VTS, hace de las diferentes situaciones que debe afrontar. El sujeto no solo debe considerar las actuaciones propias sino, además, anticipar y comprender la de aquellos buques que monitoriza dentro de la ventana espacio-temporal de referencia. A diferencia del control aéreo, los VTS carecen de capacidades ejecutivas, pudiendo solamente recomendar actuaciones a los buques, lo que limita en gran medida su capacidad de ac-

tuación. A este condicionante debe añadirse que el trabajo se desarrolla únicamente en el plano horizontal correspondiente a la superficie del mar y que los elementos objeto de control son muy heterogéneos. Todos estos son factores que el operador de un VTS debe tener en cuenta en el desempeño de su tarea, la cual contempla actuaciones que abarcan desde la previsión de una situación de peligro hasta la ordenación del tráfico marítimo. Otro elemento condicionante a considerar es la normativa que rige, a nivel local e internacional, las actuaciones tanto del operador de VTS como de los buques en la zona. El conocimiento profundo de la misma permitirá predecir los movimientos de estos últimos.

Pero más allá de lo anteriormente expuesto, el operador avezado debe poder distinguir las posibles pautas de comportamiento de los buques a partir de la información suministrada por elementos tan sencillos como el nombre, la bandera o el tipo de buque.

Los estímulos dominantes que recibe un operador VTS son visuales y auditivos. Casi toda la información con la que ha de desarrollar sus funciones va a proceder de las pantallas radar y de los receptores de radio que, sintonizados en varias frecuencias, debe monitorizar de forma simultánea.

La importancia de los estímulos auditivos se entiende mejor si se tiene en cuenta que las comunicaciones son un elemento primordial en el desempeño de la tarea del operador de VTS ya que éste es, prácticamente, su único medio de interacción con el entorno. Un operador no puede por sí mismo actuar directamente para evitar una colisión (como sí puede hacer el oficial encargado de la guardia de navegación a bordo de un barco), sino que debe esperar a que sus sugerencias y recomendaciones sean atendidas y tenidas en cuenta por los oficiales de guardia, de manera que ejerce su función a través de las acciones de otros. Esto implica que debe ser extremadamente cuidadoso tanto en la toma de decisiones como en su transmisión.

Método

La prueba, que hemos denominado Situation Awareness Test for Vessel Traffic Services [Prueba de Conciencia Situacional para Servicios de Tráfico Marítimo], en adelante SAT-VTS, se desarrolla en base al análisis de las respuestas facilitadas por los sujetos a las situaciones que se reflejan en un escenario ideal creado por miembros del equipo de investigación, que recrea la visión de la información radar obtenida por los sistemas de detección de un servicio de tráfico marítimo.

Para su construcción se contó con el soporte informático del sistema de teledetección de una estación radar real en un VTS. Con esta herramienta informática se recreó un entorno marítimo sobre el que se simulaban diversas situaciones de interacción entre buques. Una vez creado el escenario, se editó y montó en formato de video para que pudiera utilizarse tanto con PC convencionales como con sistemas multimedia estándar.

La validez es uno de los puntos esenciales en la construcción de un test, ya que "requiere comprobar la utilidad de la medida realizada, es decir, el significado de las puntuaciones obtenidas" (Macías, 2007); así mismo, la American Psychological Association (APA) y la American Educational Research Association (AERA) (AERA, APA y NCME, 1985, 2012) integran la validez predictiva y concurrente en la de criterio; además en los estándares de 1985 (AERA y APA, 2005; AERA, APA y NCME, 1985) establecen que los distintos tipos de validez (de criterio, de contenido, concurrente, etc.) tradicionalmente mencionados en la literatura son distintas formas de expresión de la validez de constructo y que debe establecerse la validez de las puntuaciones de un test en cada uso que se haga del mismo (Macías, 2007). La validez, por lo tanto, es un proceso unitario y se basa en la acumulación de evidencias (Elosua Oliden, 2003); es importante significar que no se valida el instrumento, sino las inferencias de las puntuaciones y el significado de las mismas (Messick, 1998).

De este modo se aborda el estudio de validez de la prueba desde una perspectiva global y de acumulación de evidencias: consulta a expertos, estructura interna (dimensionalidad), fundamentación teórica, relaciones con otras pruebas y consecuencias del uso del test (Messick, 1995).

Validez interna

La validez de contenido se obtiene mediante la evaluación por parte de un grupo de expertos de la pertinencia y suficiencia de los ítems (Prieto y Delgado, 2010). Para ello la prueba fue depurada y validada por 15 sujetos, operadores de VTS con una experiencia en el campo de actuación superior a 5 años ($M = 14.7$ años de experiencia en VTS), procedentes de los VTS de Algeciras, Cádiz y Tarifa, que actuaron en calidad de jueces expertos.

De igual forma se valida internamente el estudio y la construcción de los ítems, el tipo de tarea y la pertinencia del propio test (Elosua Oliden, 2003), incluyéndose el tipo de tarea, la situación y las instrucciones para la administración y corrección de la prueba. Para la realización de la corrección de la prueba se provee de un protocolo exhaustivo de administración y corrección.

Dada la dificultad de analizar un constructo psicológico con ítems no semánticos, la evidencia sobre la estructura interna se basa en un estudio de dimensionalidad de los ítems y su relación con la concepción de los mismos.

Validez externa

La evidencia basada en otras variables incluye la validez convergente y discriminante, al comparar las puntuaciones obtenidas por un grupo de sujetos que fueron sometidos a la prueba y seguidamente fueron evaluados en el único simulador certificado por IALA (International Association of Marine Aids to Navigation and Lighthouses Authorities) existente en España (situado en el Centro Jovellanos de SASEMAR, Principado de Asturias), por instructores certificados por IALA. Del mismo modo, las evidencias basadas en las consecuencias de la medición valoran las implicaciones de las puntuaciones como una base de acción futura (Pardo, 2006), realizando la comparación del rendimiento de los participantes en SAT-VTS y posteriormente en simulador, de cara a un posible desempeño en un VTS real.

El estudio de las relaciones con otras variables incluye una prueba de dígitos del WAIS (Wechsler, 1997) tanto a jueces como a *outliers* para estudiar el peso de la variable memoria en la prueba. Al resto de la muestra de alumnos no se le aplica, al considerarse que la variable se distribuye normalmente. Esta prueba se ha realizado mediante una adaptación informática, restringida a la modalidad visual, de forma que su presentación y ejecución coincidan con las de la prueba. También se aplica a este grupo la prueba de matrices progresivas de Raven para explorar la posibilidad de la influencia del CI en la puntuación en la prueba.

Consecuencias

En el presente estudio se enfatiza la importancia de la validez derivada del uso de la prueba (Messick, 1998), máxime en este caso donde se propone la construcción de una prueba específica para selección de personal. Para ello se utilizará la comparación entre el rendimiento en la prueba y las evaluaciones obtenidas en el curso VTS de IALA realizado en el Centro Jovellanos.

Participantes

En la fase final de la prueba participaron un total de 134 sujetos, alumnos de los diferentes cursos de grado y licenciatura de la Escuela de Náutica de la Universidad de Cádiz así como alumnos en prác-

ticas en buques mercantes y oficiales en activo, potenciales candidatos a operadores VTS ($M = 27.45$ años de edad, $SD = 10.79$). En la fase desarrollada mediante el uso de un simulador participaron 38 sujetos (edad, $M = 36.0$, $SD = 7.0$ años).

Diseño de la prueba

Diseño general. La herramienta desarrollada debía tener características de validez y fiabilidad psicométricas, sencillez de aplicación y corrección, economía de recursos, portabilidad y capacidad de aplicación a muestras grandes simultáneamente.

Para desarrollar la estructura del test se han tenido en cuenta, fundamentalmente, los resultados de los trabajos de Endsley, Selcon, Hardiman y Croft (1998) y, en menor medida, en los de Wiersma (2010).

La técnica de evaluación desarrollada es una adaptación del sistema de “congelado”, empleado por Endsley et al. (1998) en su test Situation Awareness Global Assessment Technique (SAGAT): después de un determinado periodo de tiempo, cuya duración el sujeto evaluado desconoce, se produce un “fundido a negro” de la pantalla de una duración determinada (y distinta en cada escenario de forma decreciente), periodo de tiempo durante el cual el sujeto evaluado deberá señalar en una carta en blanco de la zona marítima de referencia la máxima información que recuerde en relación con los buques en tránsito por la zona de estudio. También deberá señalar el rumbo y velocidad de forma vectorial, haciendo especial hincapié en las situaciones que comprometan o puedan llegar a comprometer la seguridad de estos. Se busca la sobrecarga de la memoria de trabajo, de manera que el sujeto evaluado debe necesariamente discriminar aquellas situaciones en las que verdaderamente debe fijar su atención y que debe retener, de las que no. De esta manera se garantiza la evaluación de los 3 niveles de SA, incluida la proyección (nivel 3 SA). La duración total del test es de unos 35 minutos, con un total de cuatro paradas. Estas paradas conforman los llamados “escenarios”. Estos escenarios se utilizarán posteriormente para evaluar la prueba mediante la comparación de las respuestas facilitadas por los sujetos con las plantillas de corrección.

Diseño de los ítems. Los ítems valorados se corresponden con situaciones específicas ocupadas por los buques en un momento determinado. Estos se puntúan por orden de importancia, de manera que los más importantes (aquellas situaciones en la que la seguridad de los buques está en peligro o puede estarlo) valen el doble (diez puntos) que los correspondientes a aquellas situaciones que, siendo comprometidas, no implican una amenaza para la seguridad de los buques y diez veces más que los ítems correspondientes a situaciones concebidas como meros estímulos irrelevantes que buscan sobrecargar la memoria de trabajo.

Los ítems se numeran según el escenario (1, 2, 3, 4), seguidos de una letra en orden de importancia, con mayúsculas los más importantes y minúsculas los de menor importancia (1A, 1B, 1c, 1d, etc.). Los irrelevantes no se numeran individualmente, puntuándose todos juntos.

La evaluación de los tres niveles de SA se consigue mediante la asignación y construcción de los ítems en el simulador VTS.

Los nombrados con mayúscula y que puntúan sobre 10 son las situaciones más difíciles de resolver para un operador y en las que debe poner en juego su capacidad de proyección de estatus futuro, es decir, donde tiene que retener más cantidad de información, interpretarla en función de sus conocimientos y prever que puede existir una situación de riesgo futuro. En una situación VTS real serían las situaciones que el operador debe vigilar de cerca, aunque en el momento presente no exista riesgo evidente.

Los ítems nombrados con minúscula y que se valoran sobre 5 implican que la situación de peligro es actual y obvia (y por lo tanto no entra en juego la capacidad de proyección) o bien que el buque no

presenta riesgo potencial alguno. Se pretende valorar si el sujeto está integrando la información percibida con los conocimientos que posee y su experiencia (nivel 2 SA). En un VTS real serían buques a los que el operador apenas presta atención, más que un seguimiento a intervalos.

Los ítems que no se nombran actúan tanto como estímulos distractores de los importantes como evaluadores del nivel 1 SA, que es la mera percepción de la situación, en este caso una tarea de memoria visual. Se ha pretendido que su número sea suficientemente alto y que se encuentren dispersos por todo el escenario, de forma que un operador inexperto intentará retenerlos todos en memoria, en contraposición con el más experimentado que simplemente los ignorará para no dejar de atender a los que sabe que son importantes. Esta sería la forma de actuar de un operador experto en un VTS real.

El test se basa en la valoración de la respuesta de los sujetos a los estímulos visuales, ya que se considera que es esta la fuente primordial de adquisición de información en los entornos controlados por los servicios de tráfico marítimo. Esta característica permite que sea fácilmente corregido mediante comparación con una plantilla, simplemente cotejando las respuestas facilitadas por el sujeto en su carta con la hoja de corrección.

Se ha considerado fundamental eliminar posibles desviaciones subjetivas, por lo que la zona geográfica de simulación no existe en la realidad; de esta manera se eliminan posibles ventajas por familiaridad. Adicionalmente, y aunque el test cuenta con un protocolo sencillo y exhaustivo de administración y corrección que lo hace apto para ser usado por cualquier operador experto, en esta fase de la investigación se ha considerado necesario que la corrección de las pruebas se realice siempre por la misma persona.

La prueba se ha creado con un nivel de dificultad creciente desde el primero al tercer escenario, de modo que el primer escenario sea el más fácil, el segundo más difícil, el tercero el más complicado y finalmente bajar un poco en complejidad en el cuarto y último, donde además se cuenta con el efecto del cansancio. De esta manera se pretende actuar sobre la fatiga del evaluado, buscando siempre la sobrecarga de la memoria para que los procesos de selección y toma de decisiones se mantengan activados de forma continua, ya que los sujetos desconocen cuándo se va a detener la simulación. Así se asegura la evaluación de los tres niveles de la SA, evitando que sea una prueba meramente memorística.

Variables de estudio. Las variables objeto de estudio fueron: experiencia en VTS (en años, $M = 1.95$, $SD = 4.59$), experiencia en navegación (en años, sólo si se han realizado tareas propiamente de navegación, $M = 2.85$, $SD = 5.02$), sexo (116 hombres y 32 mujeres), edad (años, $M = 27.79$, $SD = 12.12$), titulación académica (capitán del plan de estudios antiguo [anterior a 1973], licenciado, diplomado en Marina Mercante o bien grado en Navegación) y titulación profesional (capitán, piloto de 1ª ó 2ª de Marina Mercante, alumno o patrón). Las variables de salida han sido la puntuación total en el test y las puntuaciones parciales de cada escenario (4 escenarios). Igualmente se han considerado las valoraciones otorgadas por los profesores del curso VTS tanto en simulador como en el global del curso, así como las obtenidas en la prueba de dígitos del WAIS (modalidad visual), en su caso, y la puntuación en matrices progresivas de Raven (en percentiles).

Suponemos que existen variables de índole aptitudinal y actitudinal, entre las que se encuentran la memoria, la capacidad de toma de decisiones, la rapidez en el procesamiento de la información, etc., que conforman la conciencia situacional. No podemos ignorar tampoco la influencia de otras variables como motivación, cansancio, hora de día, etc., que asumiremos como parte inevitable del proceso de obtención de datos y que tendrán su reflejo en la varianza de los mismos.

Para el análisis de los resultados obtenidos por el grupo de jueces expertos y la posterior validación de la prueba se utilizó el software IBM SPSS versión 21.

Resultados

Validez interna

Prueba de jueces. En una primera fase el test fue sometido a prueba mediante grupo de jueces expertos, tras lo cual se les pasó un cuestionario para depuración de la prueba. Se solicitó a los jueces expertos información sobre los ítems y escenarios que consideraban más relevantes, puntuándose mediante una escala tipo Likert de 5 niveles. Se descartaron aquellos ítems sobre cuya relevancia al menos el 75% de los jueces no estuvieran de acuerdo, además de una escala de valoración global (para el total de la prueba). Durante el desarrollo de esta fase de diseño la prueba sufrió diversas modificaciones de poca relevancia hasta el desarrollo de la versión definitiva (versión 6.0) que se utilizó con los grupos de alumnos y pilotos sin experiencia en VTS.

El contenido y la apariencia de la herramienta quedaron validados en base a la experiencia profesional en control de tráfico marítimo de los diseñadores de la misma. Esta validación queda avalada por el criterio expresado por el grupo de jueces expertos que se muestran unánimes en cuanto a la consideración de la utilidad de la prueba para detectar candidatos idóneos para el puesto de operador VTS. El índice de validez de Aiken para el total de la prueba, obtenido tras el proceso de depuración en la fase de construcción, es de $V = .82$. Para el cálculo de este valor se ha seguido el método score (Penfield y Miller, 2004), al considerarse el más fiable y restrictivo, usando la hoja de cálculo del paquete LibreOffice. La tabla 1 muestra un resumen de varios estadísticos de los ítems.

Los jueces se muestran de acuerdo en considerar la prueba como difícil o muy difícil ($M_1 = 6.67$, $M_2 = 8.88$, dificultad para un operador y dificultad para un candidato, respectivamente) incluso para operadores expertos. En cuanto a los escenarios, existe coincidencia entre los jueces en considerar el 1º el más fácil y el 3º el más difícil. Consideran la duración del test como adecuada o dentro de la media en cuanto a longitud apreciada ($M = 5.5$ en una escala de 10).

La puntuación total tiene como media $M = 92.46$ y $SD = 48.08$.

La media en las puntuaciones totales del grupo de jueces es $M = 160.27$ y $SD = 21.59$; en el grupo de inexpertos fue de $M = 84.81$ y $SD = 44.09$. El ANOVA muestra que las diferencias de medias son significativas a más del 99.9%, $F = 42.58$.

Estructura interna

Se realiza un análisis factorial confirmatorio para apoyar nuestra hipótesis de la existencia de tres niveles definidos del constructo. La medida de adecuación muestral de KMO arroja un valor de .798, que puede ser considerado meritorio, lo que permite considerar la posibilidad de factorizar esta matriz. El test de esfericidad de Barlett tiene un nivel de significancia de .000, con un valor de $\chi^2 = 1525.233$ y 561 grados de libertad, siendo posible por tanto la factorización.

Se realiza un análisis factorial confirmatorio de tres componentes con rotación Varimax. Como se muestra en la tabla 2, el porcentaje de varianza explicada por el factor dominante sólo representa el 20.504%, valor insuficiente para sustentar la unidimensionalidad de la escala (Carmines y Zeller, 1979); sin embargo, si se considera el criterio menos restrictivo de Hattie (1984) –basado en las relaciones entre los 3 primeros factores– con un factor 5 se cumple con suficiencia (el mínimo es 3). Igualmente cumple optando por el criterio de Reckase (1979) –exige un mínimo del 20% al primer factor. La varianza explicada por el segundo y tercer factores es de 8.736% y 5.540% respectivamente, siendo el total de varianza explicada por los tres primeros factores del 34.780 %. La tabla 2 muestra simplificada la extracción de factores.

La matriz sin rotar muestra 10 factores con autovalores superiores a 1, siendo la varianza explicada por el resto de factores con autovalor > 1 del 26.575%.

Tabla 1
Estadísticos de los ítems

	Índice V de Aiken mediante método score	Correlación total de elementos corregida	Índice de dificultad
ÍTEM 1A	.91	.332	.61
ÍTEM 1B	.87	.657	.62
ítem 1c	.78	.415	.58
ítem 1d	.81	.292	.44
ítem 1e	.85	.200	.68
ítem 1f	.77	.345	.61
ítem puntos extra1	.91	.471	.62
ÍTEM 2A	.90	.564	.61
ÍTEM 2B	.80	.401	.55
ítem 2c	.74	.190	.30
ítem 2d	.81	.240	.56
ítem 2e	.82	.380	.75
ítem 2f	.75	.487	.62
ítem 2g	.77	.302	.79
ítem 2h	.78	.209	.84
ítem 2i	.76	.481	.75
ítem puntos extra2	.77	.512	.53
ítem 3A	.87	.523	.36
ítem 3B	.89	.351	.38
ítem 3C	.88	.563	.51
ítem 3d	.73	.197	.26
ítem 3e	.85	.329	.44
ítem 3f	.83	.306	.38
ítem 3g	.84	.392	.45
ítem puntos extra3	.76	.499	.54
ítem 4A	.88	.472	.52
ítem 4B	.91	.362	.51
ítem 4C	.93	.546	.41
ítem 4D	.88	.310	.34
ítem 4e	.76	.314	.21
ítem 4f	.75	.344	.79
ítem 4g	.82	.114	.20
ítem 4h	.81	.270	.57
ítem puntos extra4	.75	.468	.75

Salvo el 4D, todos los ítems importantes saturan en el primer factor, los irrelevantes o de menor importancia en el segundo, estando los demás ítems (de importancia media) repartidos entre el resto de factores, con clara preferencia por el segundo y tercero, donde no aparece ningún ítem considerado distractor.

Validez externa

Validez convergente y discriminante. Al estimar la regresión lineal simple de la puntuación en SAT-VTS sobre la puntuación otorgada por los instructores en simulador, se prueba la hipótesis de que ambas se encuentran positivamente relacionadas. El ratio $F(1, 36) = 22.287, p < .001$ es estadísticamente significativo y por tanto confirma la hipótesis de relación lineal entre ambas puntuaciones. El valor de R^2 de .382 indica que aproximadamente el 40% de la variabilidad de la puntuación en el simulador puede ser explicada por la obtenida en

Tabla 2
Estructura factorial de SAT-VTS (rotación Varimax)

	Componente		
	1	2	3
ÍTEM 1B	.805		
ÍTEM 2A	.715		
ítem 3A	.652		
ítem 1f	.582		
ítem 2e	.558		
ítem 4B	.552		
ítem 3B	.549		
ítem 2f	.539		
ÍTEM 2B	.493		
ítem 4A	.487		
ítem 3C	.477		
ítem 1c	.458		
ítem 4C	.456		
ÍTEM 1A	.346		
ítem 1e	.267		
ítem 4f	.260		
ítem puntos extra2		.763	
ítem puntos extra1		.616	
ítem puntos extra3		.608	
ítem 3d		.525	
ítem 2i		.516	
ítem puntos extra4		.502	
ítem 2c		.467	
ítem 2g		.464	
ítem 2h		.441	
ítem 4D		.424	
ítem 1d		.384	
ítem 3e			.723
ítem 3g			.546
ítem 3f			.457
ítem 4h			.433
ítem 4e			.421
ítem 2d			.336
ítem 4g			.278
Autovalores iniciales	6.971	2.970	1.883
% Varianza	20.504	8.736	5.540
% Acumulado de varianza	20.504	29.240	34.780

SAT-VTS. El análisis de residuos confirma la bondad de los datos utilizados, que se ajustan a los supuestos del modelo de regresión lineal.

Igualmente, las correlaciones entre las puntuaciones obtenidas en el test y las valoraciones finales en el total del curso VTS homologado por IALA se encuentran positivamente relacionadas, el valor de $F(1, 36) = 10.103, p < .003$, estadísticamente significativa, permite validar la hipótesis de relación lineal entre ambas puntuaciones. El valor de $R^2(.219)$ indica que aproximadamente el 22% de la variabilidad de la puntuación en el global del curso VTS puede ser explicado por el obtenido en SAT-VTS. En este caso también el análisis de residuos muestra que los datos utilizados se ajustan bien a los supuestos del modelo de regresión lineal.

Relaciones con otras variables

Un análisis de varianza entre las puntuaciones en dígitos WAIS en función del grupo de pertenencia (jueces, $M = 7.2$, $SD = 1.06$ y *outliers*, $M = 7.02$, $SD = 0.84$) muestra diferencias no significativas, $F(1, 119) = 0.117$, $p > .05$. Tampoco arroja diferencias significativas la prueba no paramétrica de Wilcoxon de las diferencias de las medianas (.102, $p > .05$).

Igualmente las puntuaciones de este grupo en matrices progresivas de Raven presentan una media de $M = 87.00$, $SD = 6.708$. Todos los *outliers* excepto uno han sido clasificados como percentil 90, grado II+, con CI superior a la media en esta escala; el resto ha sido clasificado percentil 75, grado II, superior a la media en la escala de Raven.

Fiabilidad

Estructura interna. Respecto a fiabilidad, el alfa de Cronbach es $\alpha = .864$, mostrando todos los ítems correlaciones positivas con la escala. Los ítems presentan una puntuación media de $M = 118.26$, $SD = 36.418$, para un total $N = 34$ ítems.

Los índices de discriminación (ID) de los ítems muestran valores $d \geq .30$ en todos los casos, salvo los ítems 1d, 1e, 2c, 2d, 2h, 3d, 4g y 4h. Todos ítems importantes presentan ID $> .30$, siendo el máximo el 1B (.657). El ítem de menor ID es el 4g (.114). Como podemos observar en la tabla 1, los índices de dificultad de los ítems fluctúan entre .20 y .84, existiendo por tanto una buena dispersión en la dificultad de los mismos.

En cuanto a los índices de dificultad (id) de los ítems, presentan valores consistentemente mayores en los escenarios 3º y 4º, siendo los ítems más difíciles en orden ascendente los 4e, 3d, 2c, 3A y 3B y dentro de los ítems importantes el 3B, 4D y 3A, todos ellos con valores $< .4$. El ítem más fácil de la escala es el 2h, con un índice de dificultad de .88; dentro de los ítems de mayor importancia el más fácil es el 1B con un id = .59 y el más difícil el 4D con .34.

El análisis de correlaciones de Pearson entre las variables independientes y la puntuación total en el test arroja valores significativos en $p = .01$ (bilateral) para los siguientes factores: años de experiencia VTS $r = .514$, años de navegación $r = .466$, edad $r = .488$, titulación profesional $r = .458$ y titulación académica $r = .528$; todas las variables excepto sexo presentan correlaciones significativas entre ellas en $p = .01$.

Discusión

El objetivo del presente trabajo ha sido una prueba válida y fiable para discriminar posibles candidatos a operador VTS entre las personas con cierta preparación (mínima) en navegación marítima. La herramienta obtenida es una prueba de cribado. La validez de la prueba se ha considerado desde una perspectiva global y de acumulación de evidencias dividida, exclusivamente, a efectos prácticos.

Validez de contenido

La validez de contenido se garantiza mediante la prueba de jueces y los índices de acuerdo obtenidos. En este sentido debe destacarse que el acuerdo de los jueces ha sido unánime en relación con la valoración de la pertinencia y adecuación de la prueba al objetivo propuesto. Igualmente la propia construcción de la prueba y de los ítems y la base existente de pruebas similares en campos análogos a los VTS apoyan la relevancia y representación del constructo en el test.

Validez de constructo

Tanto Burga León (2006) como Hattie (1985) señalan que “la unidimensionalidad implica que un solo rasgo o constructo se encuentra

en la base de un conjunto de ítems”; sin embargo otros autores (Wright y Linacre, 1989) indican que en la práctica ningún instrumento puede ser perfectamente unidimensional. Por otro lado, muchos factores o rasgos como la ansiedad, la motivación, la velocidad de respuesta, etc. tienen un impacto en el rendimiento de una persona en un conjunto de ítems (Hambleton, Swaminathan y Rogers, 1991). Tal como señalan Embretson y Reise (2000), lo importante es que un instrumento de medida represente un solo factor dominante, de manera que la mayor cantidad de varianza observada sea explicada por un solo atributo latente, en nuestro caso la conciencia situacional considerando sus tres niveles o estados.

Existe una clara dificultad para estudiar la estructura factorial de ítems no semánticos. Ya Endsley (1995) señalaba la necesidad de insistir en la investigación sobre validez del constructo. De esta forma, la interpretación de los resultados debe estar basada en el diseño de los ítems. Como se aprecia en la tabla 2, los ítems importantes, sin excepción, se sitúan en el primer componente de la matriz factorial, lo que puede ser interpretado como correspondiente al nivel 3 SA (proyección) siguiendo la lógica del diseño. Este primer componente se considera el principal en la estructura factorial de la prueba, representando un 20.504% de la varianza total, siendo el que más importancia tiene en la medición de la SA. En el segundo componente saturan claramente todos los ítems considerados irrelevantes y, siguiendo la lógica de la construcción de los mismos, se puede afirmar que este factor equivaldría al nivel 1 SA (percepción). El tercer factor, donde solo se encuentran ítems de media importancia, puede estar asociado con el nivel 2 SA (comprensión). En cualquier caso, todo ello debe considerarse con la debida prudencia, pues aunque el análisis indica que existe una estructura factorial, el peso total de la varianza explicada por los 3 factores considerados es bajo (un 34.780%), lo cual puede deberse a la propia naturaleza del constructo “conciencia situacional”, que agrupa, entre otros, elementos perceptivos, cognitivos, de memoria, de rapidez de procesamiento, experienciales y motivacionales. Se considera que esto queda reflejado en la estructura multifactorial de la matriz, con 10 factores con autovalores > 1 . Se puede concluir que existen evidencias suficientes sobre la validez de constructo basada en el estudio de la unidimensionalidad del instrumento, pero se hace necesario investigar con mayor profundidad.

Validez externa

Validez convergente y discriminante. El análisis de regresión refleja cómo la prueba muestra una convergencia alta con el test realizado en simulador, que es la prueba más parecida al contexto real a la que se puede someter a un candidato a operador VTS. Sin embargo, debe de considerarse que en la prueba en simulador influyen factores que no pueden ser evaluados en SAT-VTS, como son la interacción con el escenario, la información auditiva, el trabajo en equipo, etc. Sin duda estos factores son los responsables (junto con la subjetividad de los evaluadores) de gran parte de la varianza que no puede predecir el test.

También se verifica la convergencia con la nota global en el curso de operador, si bien en este caso presenta un valor más discreto. También en este caso debe tenerse en cuenta la influencia de aquellos factores que no pueden reproducirse en la prueba SAT- VTS.

Relación con otras variables

Las puntuaciones obtenidas por el grupo de jueces en la prueba de dígitos WAIS no ha mostrado diferencias significativas con las del grupo de *outliers*. En nuestra opinión, el tipo de trabajo que realizan los controladores debe influir en el entrenamiento de la memoria visual, ya que las puntuaciones han sido siempre superiores a la media en ambos grupos.

En cuanto a las puntuaciones en el test de matrices progresivas de Raven, apuntan a la existencia de altas capacidades en el grupo de *outliers* que, a falta de un análisis más completo, incide en la idea de una alta carga cognitiva y de procesamiento de información de la prueba, que premiaría a estos individuos. Esto podría asimilarse a situaciones reales que la carga de trabajo genera en los operadores VTS.

Consecuencias

La comparación entre el rendimiento en la prueba y las evaluaciones finales obtenidas en el curso VTS de IALA realizado en el Centro Jovellanos, entre las que existe una clara relación, reafirma la utilidad y la relevancia del test, que podría ser utilizado como prueba de cribado para discriminar al comienzo de una actividad formativa a los sujetos más predispuestos al aprendizaje, que pueden llevar un ritmo de trabajo superior, de aquellos otros que necesitan afianzar aptitudes básicas.

Fiabilidad

La fiabilidad de la prueba está garantizada con un muy buen valor de $\alpha = .864$ y la discriminación de los ítems es muy alta, a excepción de 8 (de un total de 34) que, aunque pueden actuar como irrelevantes, se mantienen en el test conservando su función principal de sobrecargar la memoria.

La herramienta desarrollada cumple con el objetivo propuesto de mantener un nivel de dificultad variable y creciente. Esto queda corroborado por el análisis de los índices de dificultad de los ítems, estando siempre los ítems más difíciles entre el tercer y cuarto escenarios, aportando más valor los escenarios más complejos y dando consistencia interna al test.

La prueba desarrollada excede la consideración de una prueba meramente de memoria, en primer lugar por el cuidado que se ha puesto en su diseño para generar una sobrecarga de la memoria de trabajo, ya que el sujeto evaluado desconoce el momento en que se detendrá la simulación y deberá desarrollar su mapa mental. Esto le impediría desarrollar estrategias de memorización complejas. Además, si la prueba fuera fundamentalmente una prueba de memoria los alumnos de náutica deberían alcanzar puntuaciones similares a las alcanzadas por el grupo de expertos, ya que nada hace suponer que un alumno de universidad tenga peor memoria que un operador en servicio, por mucho que éste tenga entrenada su memoria visual pues, como se sabe, la memoria de trabajo tiene limitaciones. De ello se deduce que deben existir otros procesos cognitivos que permiten seleccionar la información y desarrollar una elaboración continua de la misma, procesos que se denominan conciencia situacional y que, efectivamente, se trata de cuantificar.

Conclusión

La Unión Europea en 1999 nombró un grupo de trabajo (*task-force*) que concluyó que era necesaria más investigación sobre conciencia situacional para adaptar los nuevos VTS a los requerimientos del tráfico marítimo y la seguridad (WATERBORNE TRANSPORT Project 6.3.4, Task 27 COMFORTABLE, n.d.). De igual modo, la IALA en su último manual menciona los requisitos que se deben tener en cuenta para el entrenamiento de los operadores VTS (IALA VTS, 2008). España debe implementar los estándares de selección y formación exigidos por instancias supranacionales y que repercuten en un aumento de la calidad en la preparación y el desempeño de nuestros marinos. La necesidad de contar con una prueba de cribado para posibles candidatos a operadores en los servicios de tráfico marítimo está más que justificada. La inexistencia de herramientas psicométricas de estas características ha propiciado el desarrollo del presente trabajo.

Si bien la prueba SAT-VTS, evidentemente, presenta carencias propias de su concepción, como la limitación a una modalidad perceptiva (la visual) y la falta de interacción con el escenario, funciones cuya implementación requeriría el uso de un costosísimo simulador y la realización de pruebas individualizadas bajo supervisión de jueces expertos, sin embargo provee de una base válida y fiable para el cribado de futuros operadores en centros de control, que más adelante podrán ser evaluados en simuladores o en centros de control reales a un costo muy inferior o como herramienta de selección previa a los cursos de formación, entre otras.

El trabajo en el soporte teórico de la prueba, la conciencia situacional, proporciona un marco metodológico de análisis donde situar las funciones a evaluar e interpretar los resultados, además del necesario soporte psicológico para continuar con los trabajos en el campo del factor humano en Marina Civil, hasta ahora no tratado en la literatura en español y escasamente en otros idiomas, aunque bastante utilizado en otros campos relacionados como la navegación aérea o el control de tráfico aéreo.

De cara a un futuro próximo, se constata la necesidad de profundizar en la justificación teórica del constructo "conciencia situacional", buscando mayor soporte empírico a la validez del mismo, así como continuar evaluando y haciendo evolucionar la herramienta en contextos de selección y formación.

Extended Summary

At the beginning of the Vessel Traffic Service in Spain, with the foundation of SASEMAR (State Agency for Maritime Safety) in 1993, standard tests of attention and personality were used to select and evaluate applicants for vessel traffic services (VTS) operators, that proved to be unsatisfactory.

This project stems from the request of the SASEMAR Directorate to find a method of recruitment that fits the needs of SASEMAR, knowing that all Spanish harbors have to provide a service of maritime traffic control.

It is in the field of air traffic control and military aviation where the construct shows greater development and use. There are references to the adaptation of the concept to such varied fields as infantry combat, air combat, air traffic control, maintenance and industry, nuclear power, automotive, emergency and emergency medicine, anesthesiology, and maritime traffic control services. We have only Wiersma's (2010) dissertation.

Situation awareness. Definition and concept

Situational awareness is the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (Endsley, Farley, Jones, Midkiff, & Hansman, 1998).

Endsley's three-level theory of SA

Endsley's model is structured in three hierarchical levels of situational awareness; each level is a necessary, but not sufficient precursor below and above. This model follows a chain of information processing, beginning in perception, following through interpretation to the prediction.

Level 1 SA: Perception of environmental elements.

Level 2 SA: Comprehension of the current situation.

Level 3 SA: Prediction of future status.

Applying situation awareness to test design

In VTS, we have an overloaded work environment, with a lot of factors to consider in analyzing a situation, since the operator must not only deal with its own actions, but also to understand what the actions of the different ships that he is monitoring will be. The evaluation of the three levels of SA testing should include: measurement of working memory, comprehension of the situation (she or he must have a minimum nautical knowledge to interpret fluently one radar projection on an electronic chart), in addition to handling navigation basics and sea regulation; to assess the level 3

SA the subject must determine in what situations attention should be paid and which can be discarded, he or she must integrate all the information from the previous steps, and anticipate which vessels may become endangered.

Method

Design phase. The initial development was performed by the authors, SASEMAR operators with over 14 years in VTS.

Participants. The final phase of testing involved a total of 134 subjects, students from different graduate and undergraduate degree from the Nautical School of the University of Cadiz and trainees in merchant and seafarers in active, potential candidates to VTS operators ($M = 23.6$ years, $SD = 11.51$). Thirty-eight subjects participated in the simulator phase (Mean age = 32.0, $SD = 14.66$).

Test design

General design. Test structure is mainly based on Endsley's works (Endsley, Selcon, Hardiman, & Croft, 1998) and to a lesser extent on Wiersma's (2010).

The evaluation system is an adaptation of the "frozen" technique employed by Endsley (Endsley, Selcon, et al., 1998) in her "Situation Awareness Global Assessment Technique" (SAGAT): subjects must draw in a maritime chart of the area as much information as they remember in relation to vessels inside the area of study, with particular emphasis on (present or future) dangerous situations. These stops comprise the so-called "scenarios" (four in total), which are used for test scoring.

Item design. The items correspond to specific situations occupied by ships at any given time. Figure 2 shows two of the items. The way of evaluating the three levels of SA is accomplished by assignment and construction of the items in the VTS simulator. Items named in uppercase and scoring over 10 are the most difficult – they implied resolving situations foretelling their future status (level 3 SA). Items designated with lowercase and assessed on 5 imply either that the hazard is present and obvious or that the ship does not have any potential risk (level 2 SA). Items which are not named act both as distractors and as evaluators of Level 1 SA, the mere perception of the situation – in this case are a visual memory task.

The test was created with an increasing level of difficulty from the first to the third stage, to finally lose some complexity in the fourth, and finally the effect of fatigue has to be considered.

Test variables. The variables studied were: VTS experience (in years, $M = 1.95$, $SD = 4.59$), sailing experience ($M = 2.85$, $SD = 5.02$), sex (116 men and 32 women), age ($M = 27.79$ years, $SD = 12.12$), academic degree and professional qualifications. The output variables were the total score on the test and the partial scores of each scenario ($N = 4$). Also we have considered the ratings given by the teachers of the course both in VTS simulator and in the overall course as well as those obtained in the test of WAIS digit (visual modality), if any, and the score in Raven Progressive Matrices. We assume that there are attitudinal variables, including memory, decision making ability, speed of information processing, among others.

Results

Internal validity

Judgment by experts. The expert judges were asked to rate the items and scenarios they considered most relevant using a Likert scale of 5 levels. Items not considered relevant by at least 75% of judges were discarded. The Aiken validity index for the total test was $V = .82$. Table 1 shows summary items statistics.

The mean total scores of the group of experts was $M = 160.27$, $SD = 21.59$ and in the inexperienced group $M = 84.81$, $SD = 44.09$. The ANOVA showed that the mean differences were significant at more than 99.9%, $F = 42.58$.

Internal structure. A confirmatory factor analysis to support our hypothesis that there are three defined levels of the construct was performed. The measure of sampling adequacy, KMO, shows a value of .798, which may be considered meritorious, so factoring this matrix can be taken into account. The Bartlett's sphericity test has a significance level of .000, with value of $\chi^2 = 1525.233$ and 561 degrees of freedom, so the factorization is possible. A confirmatory factor analysis of three components with Varimax rotation was performed.

Factor analysis shows evidence of a factor solution; however, as the amount of variance explained by the dominant factor accounts for only 20.504%, this value is insufficient to support the unidimensionality of the scale (Carmines & Zeller, 1979); nevertheless, if we choose the least restrictive criteria by Hattie (1984) with a factor 5, sufficiency is met (minimum is 3). The criterion is also met if choosing Reckase's (1979). The variance explained by the second and third factors is 8.736% and 5.540% respectively, and the total variance explained by the first three factors is 34.780%. Table 2 shows a simplified extraction of factors.

External validity

Convergent and discriminant validity. Evaluated by simple linear regression of score on the SAT-VTS by score given by simulator instructors and tested the hypothesis that both are positively related. A ratio $F(1, 36) = 22.287$, $p < .001$, statistically significant, is found. The value of $R^2 = .382$, indicating that approximately 40% of the variability of the score on the simulator can be explained by the obtained by SAT-VTS. Similarly, correlations between scores on the test and endnotes in the grade approved by IALA (International Association of Marine Aids to Navigation and Lighthouses Authorities) are positively related, $F(1, 36) = 10.103$, $p = .003$, statistically significant, and therefore the hypothesized linear relationship between both scores is accomplished. The value of $R^2(.219)$ indicating that about 22% of the variability of the scores in the overall course can be explained by the obtained in SAT-VTS.

Relationships with other variables. An analysis of variance between the scores based on the WAIS digit Group (experts, $M = 7.2$, $SD = 1.06$ and outliers, $M = 7.02$, $SD = 0.84$) shows no significant differences $F(1, 119) = 0.117$, $p = .05$.

Similarly, scores of this group in Raven Progressive Matrices have a mean of $M = 87.00$, $SD = 6.708$, all outliers except one are classified as 90th percentile, grade II+, with an above average IQ on this scale. The remaining is classified 75 percentile, grade II.

Reliability

Internal structure. Cronbach's alpha is $\alpha = .864$, all items showing positive correlations with scale. The items had a mean score of $M = 118.26$, $SD = 36.418$, for a total $N = 34$ items.

Indices of discrimination (ID) of the items show d values ≥ 0.30 in all cases, except for 8 not-relevant items.

Indices of difficulty (id) of the items show values consistently greater in scenarios 3 and 4.

Discussion

The aim of this work is to create a valid and reliable test to discriminate potential candidates for VTS operator from people with some preparation (minimum) in maritime navigation. It is considered that the tool is obtained by a screening test. The validity is approached from a global perspective and accumulating evidence.

Internal validity

Content validity. Content validity is ensured by expert judgment and agreement indices obtained, noting that the agreement was unanimous in its assessment of the experts over the relevance and adequacy of the evidence to the objective.

Construct validity. There is a clear difficulty to study the factor structure of non-semantic items, and Endsley (1995) pointed out that we must insist on research on construct validity. Thus, the

interpretation of results should be based on the design of the items. It is found that the important items without exception are in the first component of the factorial matrix and it can be interpreted as corresponding to level 3 SA (projection). This first component is the main factor in the test structure, representing 20.504% of the total variance. The second component saturates all items deemed clearly distracting, and following the logic of the test building, this factor may be equivalent to Level 1 SA (perception). The third factor, whose items are only average importance, points to the level 2 SA (comprehension).

External validity

Convergent and discriminant validity. The test shows a high convergence with the test conducted in the simulator, which is most similar to the actual context of VTS. We also keep in mind that the test simulator has influencing factors that cannot be evaluated on SAT-VTS, such as interaction with the scenario, the auditory information, teamwork, and so forth. Convergence with the overall mark in the grade of operator also exists, albeit with a discrete value as above the same conclusions could be enforced.

Relationships with other variables. The scores obtained by experts on the WAIS digit test showed no significant differences from the group of outliers, which can be interpreted by the type of work operators do that can improve visual memory, as the scores have always been above average in both groups.

The scores of Raven's Progressive Matrices point to the existence of high capacities in the group of outliers, high cognitive load, and information processing work that would reward these individuals in the task.

Reliability. The reliability of the test is guaranteed with a good value of $\alpha = .864$, and item discrimination is very high, except 8 (out of 34) that may act as distractors, so they were kept in test.

Conclusion. The European Union appointed in 1999 a task force that concluded that more research was necessary in situational awareness to adapt to the new VTS requirements of maritime traffic and safety (Waterborne TRANSPORT 6.3.4 Project, Task 27 COMFORTABLE, nd) just as the IALA in its last manual lists the requirements that must be taken into account in the training of VTS operators (IALA VTS, 2008).

While SAT-VTS obviously has its own conception of shortcomings, such as being limited to a perceptual modality (visual) and the lack of interaction with the stage, it provides a valid and reliable means for screening future operators, that can be evaluated later on in simulators or in actual control centers at a much lower cost, or as a pre-screening tool training courses, inter alia.

Conflicto de intereses

Los autores de este artículo declaran que no tienen ningún conflicto de intereses.

Agradecimientos

Este trabajo se ha realizado bajo los auspicios del EPE SASEMAR (Ministerio de Fomento) en colaboración con la Universidad de Cádiz.

Referencias

- AERA, APA y NCME (1985). *Standards for Educational and Psychological Testing*.
 AERA, APA (2005) y NCME (1999). *Standards for Educational and Psychological Testing*, 146.
 Ahlstrom, U. (2007). An Eye for the Air Traffic Controller Workload. *Journal of the Transportation Research Forum*, 46(3), 103-108.
 APA (2012, September). *APA Referencing Style* (6th Edition). Washington, DC: American Psychological Association.
 Banbury, S., Tremblay, S., Rousseau, R., Forbes, K. y Breton, R. (2008). *Applying the Contextual Control Model (COCOM) to the Identification of Situation Awareness Requirements for Tactical Army Commanders*. HFES (Vol. 52, pp. 1402-1406). doi: 10.1177/154193120805201921
 Baumgartner, N., Gottesheim, W., Mitsch, S., Retschitzegger, W. y Schwinger, W. (2010). Improving Situation Awareness In Traffic Management. *Management*. Retrieved from http://www.vldb2010.org/proceedings/files/vldb_2010_workshop/QDB_2010/Paper6_Baumgartner_et_al.pdf
 Bernard, S., Ims, L., Carver, E., Cauvin, A. y Noblet, L. (2002). *The Importance of Situation Awareness in Efficiency of Distributed Collaborative Working - Application to the European Aeronautical Industry*. The 7th International Conference on Computer Supported Cooperative Work in Design (pp. 55-60).
 Blandford, A. y Wong, B. L. W. (2004). Situation awareness in emergency medical dispatch. *International Journal of Human-Computer Studies*, 61, 421-452. Retrieved from <http://discovery.ucl.ac.uk/5158/>
 BOE (2011). *Real Decreto Legislativo 2/2011, de 5 de septiembre, por el que se aprueba el Texto Refundido de la Ley de Puertos del Estado y de la Marina Mercante*.
 Bolstad, C. A., Endsley, M. R., Costello, A. M. y Howell, C. D. (2010). Evaluation of Computer-Based Situation Awareness Training for General Aviation Pilots. *International Journal of Aviation Psychology*, 20, 269-294. doi:10.1080/10508410105084142010487013
 Burga León, A. (2006). La unidimensionalidad de un instrumento de medición: perspectiva factorial. *Revista de Psicología (Lima)*, 24(1), 53-80.
 Carmine, E. G. y Zeller, R. A. (1979). *Reliability and validity assessment* (Vol. 17). Sage.
 Denford, M. J., Steele, J. A., Roy, R., y Kalantzis, E. (2004). *Measurement of air traffic control situational awareness enhancement through radar support toward operating envelope expansion of an unmanned aerial vehicle*. Proceedings of the 2004 Winter Simulation Conference (pp. 998-1006). IEEE. doi: 10.1109/WSC.2004.1371422
 Dennehy, K. C. U. (1997). *Cranfield Situation awareness Scale users manual*. Cranfield University. Retrieved from repository.tudelft.nl/assets/uuid.../Wiersma_J.W.F._PhD_Thesis.pdf
 Elosua Oliden, P. (2003). Sobre la validez de los tests. *Psicothema*, 15, 315-321.
 Embretson, S. E. y Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
 Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems: Situation Awareness. *Human Factors*, 37, 32-64. doi: 10.1518/001872095779049543
 Endsley, M. R. (1997, June). *Situation Awareness, Automation & Free Flight*. FAA/Eurocontrol Air Traffic Management R&D Seminar. Saclay, France.
 Endsley, M. R., Bolstad, C. A., Jones, D. G. y Riley, J. M. (2003, October). Situation Awareness Oriented Design: From User's Cognitive Requirements to Creating Effective Supporting Technologies. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 268-272). Denver, Colorado, USA. Retrieved from <http://www.ingentaconnect.com/content/hfes/hfproc/2003/00000047/00000003/art00004>
 Endsley, M. R., Farley, T. C., Jones, W. M., Midkiff, A. H. y Hansman, R. J. (1998). *Situation awareness information requirements for commercial airline pilots*. ICAT981 Cambridge MA Massachusetts Institute of Technology International Center for Air Transportation (p. 22). Retrieved from <http://18.729.232/handle/1721.1/35929>
 Endsley, M. R. y Robertson, M. M. (2000). Situation awareness in aircraft maintenance teams. *International Journal of Industrial Ergonomics*, 26, 301-325. doi: 10.1016/S0169-8141(99)00073-6
 Endsley, M. R., Sollenberger, R. y Stein, E. (2000). Situation awareness : A comparison of measures. *Proceedings of the Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millennium*. Savannah, GA: SA Technologies, Inc. Retrieved from <http://www.satechnologies.com/Papers/pdf/HPSAA2000-Sameas.pdf>
 Endsley, Selcon, S. J., Hardiman, T. D. y Croft, D. G. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 82-86). Santa Monica, CA: Human Factors and Ergonomics Society. Retrieved from <http://www.ingentaconnect.com/content/hfes/hfproc/1998/00000042/00000001/art00018>
 Farley, T. C., Hansman, R. J., Endsley, M. R., Amonlirdviman, K. y Vigeant-langlois, L. (1998). The Effect of Shared Information on Pilot / Controller Situation Awareness and Re-Route Negotiation. *Proceedings of the 2nd FAAEurocontrol ATM RD Seminar*.
 Hallbert, B. P. (1997). Situation awareness and operator performance: results from simulator-based studies. *Proceedings of the 1997 IEEE Sixth Conference on Human Factors and Power Plants 1997 Global Perspectives of Human Factors in Power Generation*. IEEE. doi: 10.1109/HFPP.1997.624933
 Hambleton, R. K., Swaminathan, H. y Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
 Hattie, J. (1984). An empirical study of various indices for determining unidimensionality. *Multivariate Behavioral Research*, 19, 49-78.
 Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139-164.
 Hogg, D. N., Folles, Oslash, Knut, Strand-Volden, F., Torralba, B. y Eacute (1995). Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. *Ergonomics*, 38, 2394-2413. Retrieved from <http://www.informaworld.com/10.1080/00140139508925275>
 IALA (2008). *IALA VTS MANUAL-Vessel Traffic Services Manual Edition 4*. Retrieved from [http://www.umgd.gov.pl/UserFiles/File/IALA_VTS_Manual_edition4-2008\(1\).pdf](http://www.umgd.gov.pl/UserFiles/File/IALA_VTS_Manual_edition4-2008(1).pdf)
 Itoh, M. y Inagaki, T. (1996). Design of human-interface for situation awareness. *Proceedings 5th IEEE International Workshop on Robot and Human Communication ROMAN96 TSUKUBA* (478-483). IEEE. doi:10.1109/ROMAN.1996.568884
 Javed, Y., Norris, T. y Johnston, D. (2011, May). Ontology-Based Inference to Enhance Team Situation Awareness in Emergency Management. *Proceedings of the 8th International ISCRAM Conference* (1-9). Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12517075>
 Jensen, R. S. (2013). The Boundaries of Aviation Psychology, Human Factors, Aeronautical Decision Making, Situation Awareness, and Crew Resource Management. *The International Journal of Aviation Psychology*, 259-267. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/s15327108ijap0704_1

- Jones, D., Bolstad, C., Riley, J. M., Endsley, M. R. y Shattuck, L. (2003). *Situation awareness requirements for the future objective force*. Collaborative Technology Alliances Symposium (Vol. 1).
- Kaber, D. B. y Endsley, M. R. (1998). Team situation awareness for process control safety and performance. *Process Safety Progress*, 17, 43-48. doi: 10.1002/prs.680170110
- Licao, D. L. D., Pengcheng, L. P. L. y Li, Z. L. Z. (2010). *Operator situation awareness assessment model in a nuclear power plant*, *Industrial Engineering and Engineering Management*. IEEE International Conference. Ieee. doi:10.1109/IEEM.2010.5674372
- Macías, A. B. (2007). La consulta a expertos como estrategia para la recolección de evidencias de validez basadas en el contenido. *Investigación Educativa Duranguense*, 7, 5-14.
- McNeese, M. D. (1999). Naturalistic Decision Making (Book Review). *International Journal of Cognitive Ergonomics*, 3, 373-375. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=3171494&site=ehost-live>
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5-8.
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45, 35-44.
- Pardo, C. A. (2006). *Hablemos de validez... Taller del Instituto Colombiano para el Fomento de la Educación Superior*. Retrieved from http://w3.icfes.gov.co:8080/seminariointernacional/ktmllite/files/uploads/CARLOS_PARDO_-_ICFES.pdf
- Penfield, R. D. y Miller, J. M. (2004). Improving content validation studies using an asymmetric confidence interval for the mean of expert ratings. *Applied Measurement in Education*, 17, 359-370.
- Prieto, G. y Delgado, A. (2010). Fiabilidad y validez. *Papeles Del Psicólogo*, 31, 67-74.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational and Behavioral Statistics*, 4, 207-230.
- Salmon, P. M. y Stanton, N. A. (2013). Situation awareness and safety: Contribution or confusion? Situation awareness and safety editorial. *Safety Science*, 56, 1-5. doi: 10.1016/j.ssci.2012.10.011
- Salmon, P. M., Young, K. L. y Cornelissen, M. (2013). Compatible cognition amongst road users: The compatibility of driver, motorcyclist, and cyclist situation awareness. *Safety Science*, 56, 6-17. doi: 10.1016/j.ssci.2012.02.008
- Schömig, N. y Metz, B. (2013). Three levels of situation awareness in driving with secondary tasks. *Safety Science*, 56, 44-51. doi: 10.1016/j.ssci.2012.05.029
- Stubbings, L., Chaboyer, W. y McMurray, A. (2012). Nurses' use of situation awareness in decision-making: an integrative review. *Journal of Advanced Nursing*, 68, 1443-1453. doi: 10.1111/j.1365-2648.2012.05989.x
- Uhlarik, J. y Comerford, D. A. (2002). *A review of situation awareness literature relevant to pilot surveillance functions*. DTIC Document.
- Underwood, G., Ngai, A. y Underwood, J. (2012). Driving experience and situation awareness in hazard detection. *Safety Science*, 56, 29-35. doi: 10.1016/j.ssci.2012.05.025
- Walker, G. H., Stanton, N. A. y Chowdhury, I. (2013). Self Explaining Roads and situation awareness. *Safety Science*, 56, 18-28. doi: 10.1016/j.ssci.2012.06.018
- Wechsler, D. (1997). *WAIS-III*. New York, NY: Psychological Corporation.
- Wickens, C. D. (2002). Situation Awareness and Workload in Aviation. *Current Directions in Psychological Science*, 11, 128-133. doi:10.1111/1467-8721.00184
- Wiersma, J. W. F. D. U. of T. (2010). *Assessing Vessel Traffic Service Operator Situation Awareness*. Delft. Retrieved from repository.tudelft.nl/assets/uuid.../Wiersma_J.W.F._PhD_Thesis.pdf
- Wright, B. D. y Linacre, J. M. (1989). Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation*, 70, 857.
- Young, K. L., Salmon, P. M. y Cornelissen, M. (2013). Missing links? The effects of distraction on driver situation awareness. *Safety Science*, 56, 36-43. doi: 10.1016/j.ssci.2012.11.004
- Zhang, Y., Drews, F. A., Westenskow, D. R., Foresti, S., Agutter, J., Bermudez, J. C., ... Loeb, R. (2002). Effects of integrated graphical displays on situation awareness in anaesthesiology. *Cognition Technology and Work*, 4, 82-90. doi: 10.1007/s101110200007