

Aspectos metodológicos de la evaluación subjetiva de la carga mental de trabajo

Rubio, S., Díaz, E.M. y Martín, J.

RESUMEN

Objetivo: El objetivo de este trabajo es evaluar las propiedades psicométricas de tres instrumentos de medida subjetiva de la carga mental de trabajo: NASA-TLX, SWAT y WP.

Material y métodos: La muestra estuvo formada por 36 sujetos que realizaron dos tareas experimentales, de forma separada y simultáneamente. Las dos tareas experimentales fueron una tarea de búsqueda en la memoria de Sternberg y una tarea de seguimiento (tracking). **Resultados:** Los resultados demostraron que: a) los tres instrumentos evaluados fueron sensibles a las manipulaciones realizadas en la dificultad de la tarea, revelándose la técnica WP como la mejor; b) el poder diagnóstico de WP y SWAT fue claramente superior al obtenido por TLX, y c) la validez convergente de los tres fue elevada.

PALABRAS CLAVE

Carga mental subjetiva. Sensibilidad. Validez. Poder diagnóstico.

METHODOLOGY ISSUES IN SUBJECTIVE EVALUATION OF MENTAL LOAD AT WORK

SUMMARY

Objective: The present research evaluates several psychometric properties of three multidimensional subjective workload assessment instruments: NASA-TLX, SWAT and WP.

Material and methods: Thirty six subjects performed two laboratory tasks separately and simultaneously. A discrete Sternberg memory task and a continuous tracking task were used.

Results: The results showed that: a) the three instruments were sensitive to the difficulty task manipulations, but WP was the best; b) diagnosticity of WP and SWAT was clearly superior to the obtained using TLX, and c) convergent validity of the three instruments was elevated.

KEYWORDS

Subjective mental workload. Multidimensional instruments. Sensitivity. Validity. Diagnosticity.

INTRODUCCIÓN

En la actualidad, la evaluación de la carga mental es un aspecto central en la investigación y el desarrollo de sistemas hombre-máquina que permitan obtener grados más altos de bienestar, satisfacción, eficacia y seguridad en el trabajo, objetivos fundamentales de la ergonomía.

La carga mental de trabajo se define como la diferencia entre las demandas cognitivas de un puesto de trabajo o una tarea y la capacidad de atención del trabajador¹⁻⁴. Actualmente, existe un acuerdo general en admitir que la «carga mental» es un concepto multidimensional y, por tanto, está determinado por diferentes factores o dimensiones^{2,3}. Sin embargo, el número y el tipo de dimen-

siones que determinan la carga mental todavía no están claros. Por ejemplo, Wickens⁴ distingue diversas dimensiones de carga mental en función del tipo de recursos de procesamiento que demande la tarea: estados de procesamiento perceptivo-central o de respuesta, códigos de procesamiento espacial o verbal; input auditivo o visual, y respuesta oral o manual. Hart y Staveland⁵ distinguen las siguientes 6 dimensiones: demanda mental, demanda física, demanda temporal, ejecución, esfuerzo y grado de frustración. En esta misma línea, Reid et al⁶ asumen que la carga está principalmente compuesta por tres dimensiones: carga debida al tiempo, carga debida al esfuerzo mental y carga debida al estrés. En resumen, el número de dimensiones de la carga que distinguen los autores suele estar determinado por supuestos a priori sobre las distintas fuentes de carga, más que partir del análisis estadístico de los datos. Con el tiempo, las distintas dimensiones de carga se han ido reduciendo, de manera que algunas de ellas han entrado a formar parte de una sola dimensión más general⁵. Aunque se ha propuesto un número variable de dimensiones, parece exis-

Correspondencia:

Susana Rubio Valdehita.

Departamento de Psicología Diferencial y del Trabajo. Facultad de Psicología. Universidad Complutense de Madrid. Campus de Somosaguas. 28223 Madrid. Correo electrónico: psdif48@sis.ucm.es

Trabajo recibido el 18-IV-00. Aceptado el 27-VI-01.

tir cierto acuerdo en que la carga, fundamentalmente la subjetiva, se debe a tres grandes áreas o fuentes. La primera englobaría todos los aspectos relativos a la presión temporal de la tarea (tiempo disponible, tiempo necesitado). La segunda estaría formada por variables que hacen referencia a la cantidad de recursos de procesamiento que demanda la tarea (mental, sensorial, tipo de tarea, etc.). Por último, la tercera dimensión general de carga se relacionaría con aspectos de naturaleza más emocional (fatiga, frustración, nivel de estrés, etc.).

Por otro lado, se ha propuesto una variedad de técnicas de predicción y evaluación de la carga mental. La utilidad de los diferentes procedimientos de evaluación de la carga mental es función del grado en el que éstos cumplen los siguientes criterios: sensibilidad, poder de diagnóstico, selectividad/validez, grado de intrusión, fiabilidad, requisitos de implementación y aceptación por el operador⁷. La mayoría de los métodos utilizados en la evaluación de la carga mental se pueden clasificar dentro de las tres categorías generales siguientes⁸: a) procedimientos basados en el rendimiento; b) procedimientos subjetivos, y c) medidas fisiológicas.

La utilización de medidas del rendimiento como un índice de la carga mental de trabajo se basa en el supuesto de que todo aumento en la dificultad de una tarea producirá un incremento en sus demandas, que se pondrá de manifiesto reduciendo el rendimiento. La principal ventaja de estas medidas es su elevado poder de diagnóstico.

Con el uso de indicadores fisiológicos se parte del supuesto de que la carga mental de una tarea se puede medir a través del grado de activación fisiológico. Entre sus inconvenientes destacan sus enormes requisitos de implementación, la mala aceptación que reciben por parte de los sujetos que participan en la evaluación y, lo más importante, las dudas sobre su validez como índices de la carga mental del trabajo. Entre ellos destacan la medida del componente P300 de los potenciales evocados, el diámetro pupilar y la frecuencia cardíaca.

Los procedimientos subjetivos asumen que un mayor gasto de capacidad está asociado con los sentimientos subjetivos de esfuerzo, y que éstos pueden ser evaluados adecuadamente por los individuos. Existe una gran variedad de procedimientos subjetivos que se han aplicado en la evaluación de la carga mental. Entre ellos, destacan la Escala de Cooper-Harper, la de Bedford, SWAT (Subjective Workload Assessment Technique)⁹, NASA-TLX (Task Load Index)⁵ y WP (Workload Profile)¹⁰. La gran variedad de técnicas subjetivas desarrolladas para evaluar la carga mental ha supuesto un grave inconveniente a la hora de decidirse por el uso de una de las técnicas en concreto. Esto ha llevado a diversos autores a estudiar las características de las diferentes técnicas con el objetivo de establecer una metodología en la que se reflejen las propiedades a tener en cuenta a la hora de elegir entre las distintas técnicas, en función de cuál sea el objetivo y/o el ámbito de investigación. Por sus características particulares (mínimos requisitos de implementación, elevada aceptación por los trabajadores, buena validez y fiabilidad, etc.), los instrumentos subjetivos son los utilizados

con más frecuencia en contextos aplicados. Por todo ello, la presente investigación tiene como principal objetivo estudiar y comparar las propiedades psicométricas y metodológicas de tres instrumentos multidimensionales de evaluación subjetiva de la carga mental (NASA-TLX, SWAT y WP), en función de los criterios de sensibilidad, poder de diagnóstico, validez, grado de instrucción, requisitos de implementación y aceptación del operador. Los resultados de este estudio podrán servir de ayuda a los profesionales de la psicología en la elección de la técnica de evaluación de la carga mental más adecuada para una determinada situación.

MATERIALES Y MÉTODOS

Muestra

En esta investigación participaron 36 sujetos voluntarios estudiantes de la Facultad de Psicología de la Universidad Complutense de Madrid, de edades comprendidas entre 20 y 24 años, todos ellos diestros. Los participantes fueron distribuidos aleatoriamente en tres grupos de igual tamaño: 12 contestaron al instrumento TLX, 12 al SWAT y los 12 restantes al Perfil de Carga (WP).

Tareas

Se utilizaron dos tareas experimentales: la tarea de búsqueda en la memoria de Sternberg y una tarea de seguimiento (tracking). Ambas tareas fueron aplicadas mediante un programa de ordenador especialmente diseñado para esta investigación. En la tarea de memoria de Sternberg los sujetos debían memorizar un conjunto de letras (consonantes) al comienzo de cada ensayo. Se establecieron dos niveles de dificultad objetiva según el número de letras a memorizar: la versión más fácil requería memorizar solamente 2 letras (m2) y la más difícil 4 (m4). Durante los ensayos experimentales, la tarea de los sujetos consistía en responder si la letra que se le presentaba era una de las que había memorizado previamente. Esta tarea se realizaba siempre con la mano izquierda. El rendimiento en esta tarea fue medido a través del tiempo medio de las respuestas correctas (en centésimas de segundo) para cada ensayo.

La tarea de seguimiento consistía en mantener el cursor dentro de un camino móvil de referencia, utilizando las teclas de cursor derecha-izquierda del teclado. Como parámetro de dificultad objetiva de esta tarea se manipuló el ancho de la vía: siendo el nivel 1 el más estrecho (mayor dificultad, s1) y el nivel 3 el más ancho (menor dificultad, s3). Esta tarea fue realizada con la mano derecha. Como medida del rendimiento se consideró la raíz cuadrada del error, o desviación de la vía, cuadrático medio (RMSE), calculado según la fórmula siguiente:

$$RM = \frac{D^2}{n}$$

donde D es el grado de desviación del camino de referencia (en píxeles), y n es el número de veces que el sujeto se desvía del camino.

A partir de la combinación de las tareas anteriores, se obtuvieron cuatro tareas duales (m2s1, m2s3, m4s1 y m4s3). Los sujetos recibieron la instrucción de prestar la misma atención para las dos tareas y hacer lo mejor posible ambas.

Instrumentos de evaluación de la carga mental

En esta investigación se utilizaron tres instrumentos subjetivos de evaluación de la carga mental de trabajo, que se describen a continuación.

La SWAT (Subjective Workload Assessment Technique), desarrollada por el grupo de investigación de Reid⁶, utiliza procedimientos de análisis de datos basados en las técnicas de medida conjunta (conjoint measurement). Asume que la carga mental de una tarea o actividad está determinada por tres factores o dimensiones que los autores denominan tiempo, esfuerzo mental y estrés. Cada dimensión se evalúa en una escala de tres puntos con descripciones verbales, que describe en la tabla 1.

El procedimiento de aplicación de SWAT consta de dos fases. En la primera fase, denominada «desarrollo de la escala», se combinan los tres niveles de cada una de las tres dimensiones obteniéndose 27 descripciones posibles ($3 \times 3 \times 3$) y los sujetos deben ordenar estas 27 combinaciones en función del nivel de carga que les produce cada una de ellas, antes de realizar o de conocer la tarea que se va a evaluar. Se aplica el análisis de medida conjunta a las ordenaciones dadas por los individuos para obtener una escala interválica, que asigna una puntuación de 0 a 100 a cada una de las 27 combinaciones. A continuación, los sujetos se agrupan en función de la dimensión a la que dieron más importancia en la ordenación y se obtiene una escala de carga mental

para cada grupo. En la segunda fase, denominada «fase de valoración», los sujetos evalúan la carga mental de la tarea, inmediatamente después de haberla realizado, asignando un 1, un 2 o un 3 en cada una de las tres dimensiones. Estas valoraciones se transforman en una puntuación de carga aplicando la escala desarrollada en la fase anterior.

La NASA-TLX (Task Load Index) fue desarrollado por Hart y Staveland⁵ y es el que utiliza la NASA en sus evaluaciones de la carga mental. Distingue las seis dimensiones de carga mental que se nombran en la tabla 2.

Al igual que SWAT, el procedimiento de aplicación consta de dos fases. La primera fase («de ponderación»), previa a la realización de la tarea, consiste en realizar las 15 comparaciones binarias de las 6 dimensiones, eligiendo, de cada par, la que el sujeto percibe como mayor fuente de carga. Para cada dimensión se obtiene un peso que viene dado por el número de veces que ésta haya sido seleccionada en las comparaciones binarias. Este peso puede variar entre 0 (la dimensión no ha sido elegida en ninguna de las comparaciones) y 5 (la dimensión ha sido elegida en todas las comparaciones en las que aparecía).

En la segunda fase («valoración»), inmediatamente después de realizar la tarea, el sujeto tiene que estimar, en una escala de 0 a 100 (dividida en intervalos de 5 unidades) la carga mental de dicha tarea debida a cada una de las 6 dimensiones.

Con los datos obtenidos en estas dos fases se calcula un índice global de la carga mental de la tarea aplicando la siguiente fórmula:

$$IC = \sum_{i=1}^6 p_i X_i / 15$$

donde IC es índice de carga, p_i es el peso obtenido para cada dimensión en la fase de ponderación, y X_i es la puntuación obtenida por la dimensión en la fase de valoración.

Tabla 1. Dimensiones del SWAT (Subjective Workload Assessment Technique)

Dimensiones	Puntuación	Descripción
Tiempo	1	Normalmente sobra tiempo. Las interrupciones o superposiciones entre las actividades son muy infrecuentes o nunca ocurren
	2	Ocasionalmente sobra tiempo. Las interrupciones o superposiciones entre las actividades son frecuentes
	3	Nunca o casi nunca sobra tiempo. Las interrupciones o superposiciones entre las actividades son muy frecuentes o se producen siempre
Esfuerzo mental	1	Se requiere muy poco esfuerzo o concentración mental consciente. La actividad es casi automática, y requiere muy poca o ninguna atención
	2	Se requiere un nivel moderado de esfuerzo o concentración mental consciente. La complejidad de la actividad es moderadamente alta debido a la incertidumbre, la imprevisión o la falta de familiaridad. Se requiere un nivel de atención considerable
	3	Se necesita un nivel alto de esfuerzo mental y de concentración. La actividad es muy compleja y requiere total atención
Estrés	1	Niveles muy bajos de confusión, riesgo, frustración o ansiedad, que pueden tolerarse con facilidad
	2	Se producen niveles moderados de estrés debido a confusión, frustración o ansiedad. Para mantener el nivel adecuado de rendimiento es necesario hacer un esfuerzo significativo
	3	Se producen niveles muy intensos de estrés debido a confusión, frustración o ansiedad. Se requiere un grado de autocontrol extremo

Tabla 2. Dimensiones del NASA-TLX (Task Load Index)

Dimensiones	Descripción
Demanda mental	Cantidad de actividad mental y perceptiva que requiere la tarea (p. ej., pensar, decidir, calcular, recordar, mirar, buscar, etc.)
Demanda física	Cantidad de actividad física que requiere la tarea (p. ej., pulsar, empujar, girar, deslizar, etc.)
Demanda temporal	Nivel de presión temporal sentida. Razón entre el tiempo requerido y el disponible
Rendimiento	Hasta qué punto el sujeto se siente insatisfecho con su nivel de rendimiento
Esfuerzo	Grado de esfuerzo mental y físico que tiene que realizar el sujeto para obtener su nivel de rendimiento
Nivel de frustración	Hasta qué punto el sujeto se siente inseguro, estresado, irritado, descontento, etc., durante la realización de la tarea

Recientemente, Tsang y Velázquez¹⁰, basándose en el modelo de recursos múltiples enunciado por Wickens⁴, han propuesto una técnica novedosa que intenta recoger las ventajas de los procedimientos basados en el rendimiento en situaciones de tarea dual (elevado poder de diagnóstico) y las de los procedimientos subjetivos (buena aceptación, requisitos de implementación muy escasos y nada intrusivos).

El WP (Workload Profile) se encuentra todavía en fase de desarrollo, y como las propias autoras reconocen, todavía es necesario investigar más en profundidad las propiedades de este procedimiento antes de establecer conclusiones definitivas sobre su utilidad y aplicabilidad. A pesar de ello, los resultados obtenidos hasta el momento permiten pensar que se trata de un procedimiento subjetivo bastante prometedor.

A diferencia de los dos procedimientos subjetivos anteriores, el Perfil de Carga Mental se aplica en una sola fase, posterior a la realización de las tareas. Utiliza una matriz de tantas filas como tareas y combinaciones entre ellas, y 8 columnas (una para cada tipo de recurso establecido por el modelo de Wickens). Los sujetos estiman la proporción de recursos atencionales de cada tipo utilizados en la realización de la/s tarea/s (asignando un valor de 0 a 1).

Variables

Las variables independientes consideradas en la presente investigación fueron: la dificultad de la tarea de memoria (m2-m4), la complejidad de la tarea de seguimiento (s3-s1), la situación de evaluación (tarea simple-tarea dual) y el instrumento de medida de la carga mental subjetiva utilizado (TLX-SWAT-WP). Las tres primeras variables fueron de medidas repetidas, es decir, que todos los sujetos realizaron todas las posibles combinaciones de los factores, y la última de medidas independientes, es decir, que cada uno de los tres grupos de sujetos contestó solamente a una de las escalas subjetivas.

Como variables dependientes se recogieron las medidas del rendimiento mencionadas anteriormente para cada una de las tareas y las correspondientes evaluaciones subjetivas de la carga mental propias de cada escala.

Todas las sesiones tuvieron lugar en el Laboratorio de Psicología del Trabajo de la Facultad de Psicología de la Universidad Complutense de Madrid.

RESULTADOS

Grado de intrusión

En primer lugar, y con el objetivo de analizar la existencia de diferencias en el rendimiento de los individuos debidas al instrumento de evaluación de la carga mental aplicado en cada sesión, se realizó un análisis de variancia (ANOVA) sobre las medidas del rendimiento. Como puede observarse, en la tabla 3, no se encontraron efectos significativos del instrumento aplicado sobre ninguna de las variables del rendimiento utilizadas ($p > 0,05$ en todos los casos). Teniendo en cuenta que en los tres casos se trata de técnicas de «papel y lápiz» que requieren que el sujeto responda una vez realizada la tarea experimental, es válido asumir que el grado de intrusión generado por estas escalas es prácticamente nulo.

Sensibilidad de las medidas de carga mental

La sensibilidad de las medidas de carga mental proporcionada por cada uno de los instrumentos de evaluación considerados fue analizada mediante la realización de análisis de variancia. Se trataba de investigar hasta qué punto los índices globales de carga mental diferían en función de las variaciones objetivas de la dificultad de las tareas tanto sim-

Tabla 3. Resultados del análisis de variancia (ANOVA) comparando los niveles de rendimiento para los tres grupos

Variables	F	Significación estadística
M2_tiempo	2,006	0,151
M4_tiempo	0,479	0,623
M2S3_tiempo	0,127	0,881
M4S3_tiempo	0,459	0,636
M2S1_tiempo	2,288	0,060
M4S1_tiempo	2,876	0,071
S3_rmse	1,101	0,344
S1_rmse	0,470	0,629
M2S3_rmse	1,550	0,227
M4S3_rmse	0,086	0,917
M2S1_rmse	0,906	0,414
M4S1_rmse	1,122	0,338

ples como duales, ya que un instrumento sensible es el que permite obtener estimaciones de carga mental significativamente más altas cuando la tarea es objetivamente más compleja. Los resultados obtenidos para cada uno de los instrumentos (NASA-TLX, SWAT y WP) aparecen en la tabla 4.

Validez

Para el estudio de la validez se promediaron los valores de carga mental para cada instrumento y para cada una de las tareas, y se calcularon los coeficientes de correlación de Pearson entre los índices proporcionados por los tres instrumentos. Los resultados de este análisis aparecen en la tabla 5. Todas las correlaciones fueron positivas y estadísticamente significativas ($p < 0,001$, en todos los casos), lo cual permite concluir que la validez convergente de los tres instrumentos es muy elevada.

Poder de diagnóstico

El poder de diagnóstico de cada uno de los instrumentos fue analizado mediante la realización de análisis discriminantes por el procedimiento de pasos sucesivos. El objetivo de estos análisis es determinar hasta qué punto los perfiles de carga mental sirven para discriminar entre las tareas. Los perfiles de carga de cada una de las tareas vienen dados por las valoraciones asignadas por los sujetos en cada una de las dimensiones que distinguen los instrumentos de evaluación. Por tanto, en estos análisis se introdujeron como variables discriminantes los perfiles de carga y como variable de agrupamiento la tarea, considerando tanto las tareas simples como las duales.

En primer lugar, se obtuvieron los perfiles de carga mental de cada tarea para el TLX, que aparecen en la figura 1.

Como se observa claramente en la figura 2, la primera función discriminante distingue las dos versiones de ta-

rea: simple y dual, ya que todas las tareas simples se sitúan a la izquierda del origen (valores negativos en el eje horizontal), mientras que todas las duales se sitúan en el margen derecho (valores positivos en el eje horizontal).

En la figura 3 aparecen los perfiles de carga de cada tarea y en la figura 4 se presenta el resultado del análisis y la situación de cada tarea en el espacio bidimensional obtenido por el SWAT.

La primera función discriminante, al igual que en el caso del instrumento TLX, discrimina entre las dos versiones de tarea: simple y dual. Todas las tareas simples se sitúan a la izquierda del origen (valores negativos en el eje horizontal), mientras que todas las tareas duales se sitúan en el margen derecho (valores positivos en el eje horizontal).

En este caso, la segunda función discriminante (eje vertical) presenta un mayor poder de diferenciación entre las tareas. Así, considerando las cuatro tareas simples, la segunda función discrimina claramente entre las tareas de memoria y las de seguimiento: las tareas de memoria se sitúan por encima del origen (valores positivos), en el cuadrante superior izquierdo, y las dos tareas de seguimiento en el cuadrante inferior izquierdo (valores negativos). En cuanto a las tareas duales, la combinación s3-dual obtiene valores positivos y la s1-dual valores negativos.

Como con los dos instrumentos anteriores, también para el WP, el primer paso fue la obtención de los perfiles de carga mental de cada tarea. Estos perfiles aparecen en la figura 5.

En la figura 6 aparece la situación de cada tarea en el espacio formado por las dos funciones discriminantes. Como se puede observar en la figura 6, ahora es la segunda función la que discrimina entre las dos versiones de tarea: simple y dual. Todas las tareas simples se sitúan por debajo del origen (valores negativos en el eje vertical), mientras que todas las tareas duales se sitúan en la mitad superior (valores positivos en el eje vertical).

Tabla 4. Resumen de los análisis de variancia (ANOVA) para cada uno de los índices utilizados

Tarea	Factor	Índice	F (1,11)	p
Simple	Número de letras a memorizar	TLX	1,09	0,319
		SWAT	0,00	0,976
		WP	28,72	0,000 ^a
	Ancho del camino de referencia	TLX	48,86	0,000 ^a
		SWAT	14,20	0,003 ^a
		WP	50,43	0,000 ^a
Dual	Número de letras a memorizar (M)	TLX	13,81	0,003 ^a
		SWAT	4,52	0,057
		WP	70,39	0,000 ^a
	Ancho del camino de referencia (A)	TLX	43,67	0,000 ^a
		SWAT	18,14	0,001 ^a
		WP	30,93	0,000 ^a
	Interacción M × A	TLX	4,86	0,500
		SWAT	0,03	0,864
		WP	6,22	0,030 ^b

^a $p < 0,01$. ^b $p < 0,05$.

TLX: Task Load Index; SWAT: Subjective Workload Assessment Technique; WP: Workload Profile.

Tabla 5. Coeficientes de correlación de Pearson entre los tres índices globales de carga mental

	TLX	SWAT	WP
TLX	1,0000		
SWAT	0,9817	1,0000	
WP	0,9863	0,9720	1,0000

TLX: Task Load Index; SWAT: Subjective Workload Assessment Technique; WP: Workload Profile.

En cuanto a la condición de tarea simple, la primera función discriminante (eje horizontal) sirve para diferenciar las tareas de memoria (valores negativos en el eje horizontal) de las tareas de seguimiento (valores positivos en el eje horizontal).

En relación con la situación de tarea dual, la diferenciación entre las cuatro tareas a partir de la primera función discriminante es menor, si bien la tarea dual más compleja (m4s1) es la única que obtiene valores positivos en esta dimensión (eje horizontal), mientras que los valores para el resto de tareas duales son negativos.

DISCUSIÓN

El análisis de la carga mental que produce un trabajo tiene implicaciones prácticas muy diversas en el estable-

cimiento de planes de formación y entrenamiento de los trabajadores, en el proceso de selección de personal, y en el diseño y rediseño de sistemas y tareas.

Tener que procesar diversas informaciones al mismo tiempo, tomar las decisiones adecuadas, resolver los problemas de manera eficaz en situaciones de emergencia y la inevitable necesidad de adaptarse a los cambios producidos por los avances tecnológicos han producido un aumento considerable en la complejidad cognitiva de un gran número de las actividades. Esto hace que un objetivo prioritario de la psicología sea el análisis de las demandas que imponen las diferentes tareas, con el fin de diseñar puestos de trabajo que produzcan una menor carga mental. Esta reducción en los niveles de carga mental se traduce en niveles más bajos de estrés y de accidentalidad y en una reducción de la probabilidad de cometer errores.

La presente investigación pretendía analizar y comparar las propiedades de tres instrumentos de medida de la carga mental subjetiva: NASA-TLX, SWAT y WP. En este sentido, se analizó el grado en el que cada uno de estos instrumentos poseía las siguientes propiedades: grado de intrusión, sensibilidad, validez, poder de diagnóstico, requisitos de implementación y aceptación por parte de los sujetos.

Como era previsible, no se encontraron diferencias en el rendimiento debidas al instrumento de evaluación de la carga mental utilizado. Por tanto, y teniendo en cuenta que la aplicación de estas técnicas es posterior a la realización de la tarea que está siendo evaluada, podemos concluir que el grado de intrusión o de interferencia con el

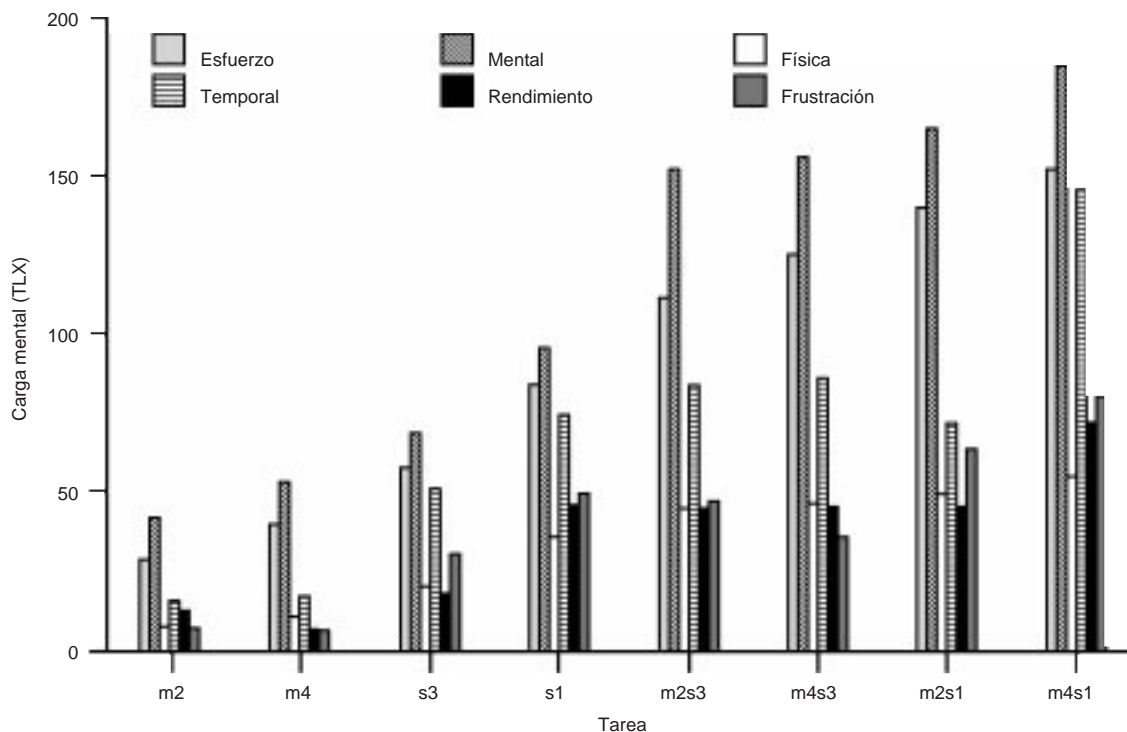


Figura 1. Perfiles de carga mental en TLX (Task Load Index).

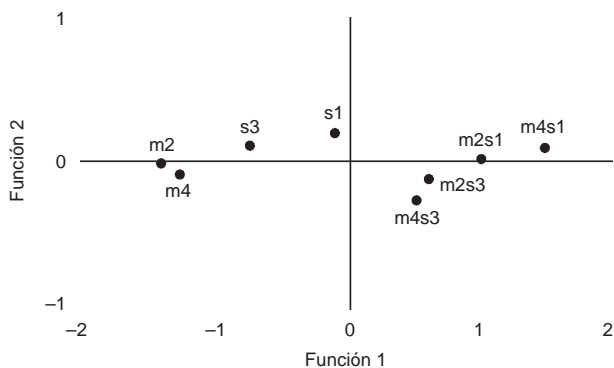


Figura 2. Resultado del análisis discriminante para cada tarea en el instrumento TLX.

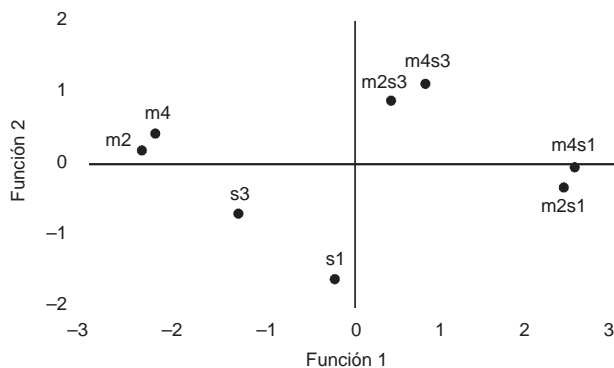


Figura 4. Resultados del análisis discriminante en el instrumento SWAT.

rendimiento de estos tres procedimientos es inapreciable. Como ya se ha mencionado, el escaso grado de intrusión que presentan la gran mayoría de los procedimientos subjetivos de evaluación de la carga mental representa una de las principales ventajas de los mismos, frente a otros indicadores de tipo fisiológico o medidas del rendimiento^{2,4}.

Aunque los tres instrumentos empleados para medir la carga mental proporcionaron índices globales de carga que resultaron sensibles a las variaciones en la dificultad objetiva de las tareas, el Perfil de Carga (WP) fue el único en el que se pusieron de manifiesto las diferencias causadas por los dos factores de complejidad de las tareas y por la interacción entre ambos. Por tanto, podemos concluir que de los tres instrumentos considerados, el WP es el que presenta una mayor sensibilidad, la cual también se

puso de manifiesto en el estudio de Tsang y Velázquez¹⁰.

En cuanto a los otros dos instrumentos, ambos demostraron una sensibilidad muy similar entre sí. Sin embargo, el NASA-TLX puede considerarse ligeramente más sensible si tenemos en cuenta el tamaño de los valores obtenidos por el estadístico F. Este resultado coincide con las ideas expresadas por autores como Nygren¹¹.

En cuanto a la validez, se trataba fundamentalmente de comprobar la validez convergente del instrumento más novedoso, el WP, en función del grado de relación que presentara con los otros dos instrumentos, cuya validez está sobradamente demostrada^{8,11}. En la presente investigación, se obtuvieron coeficientes de correlación positivos y muy cercanos a la unidad, lo cual demuestra la elevada validez convergente de los tres instrumentos.

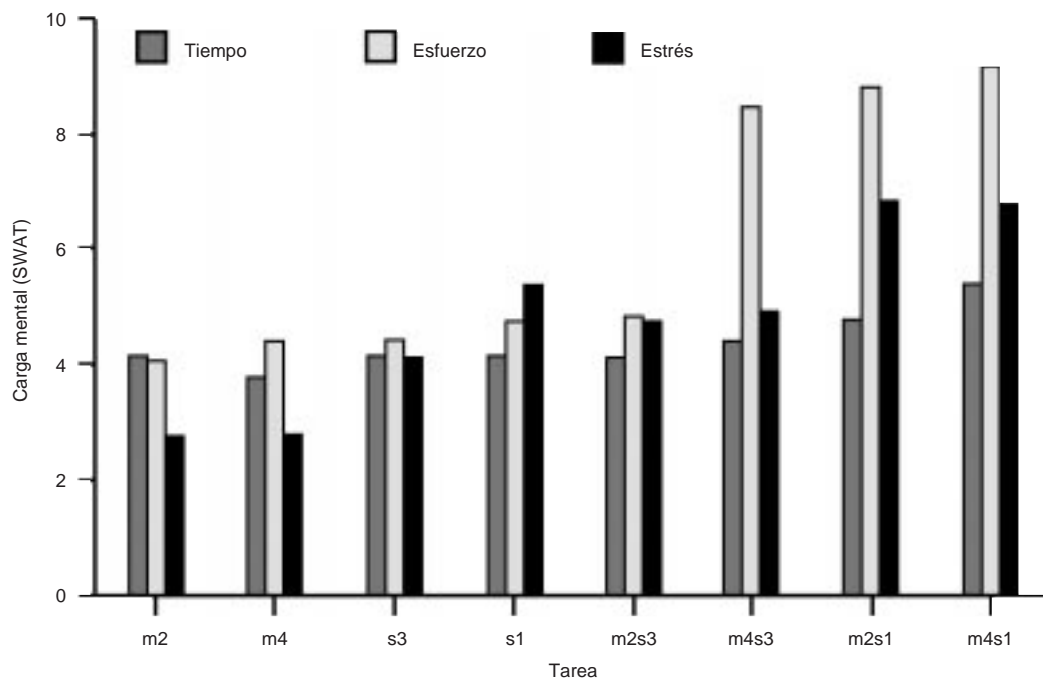


Figura 3. Perfiles de carga mental en SWAT (Subjective Workload Assessment Technique).

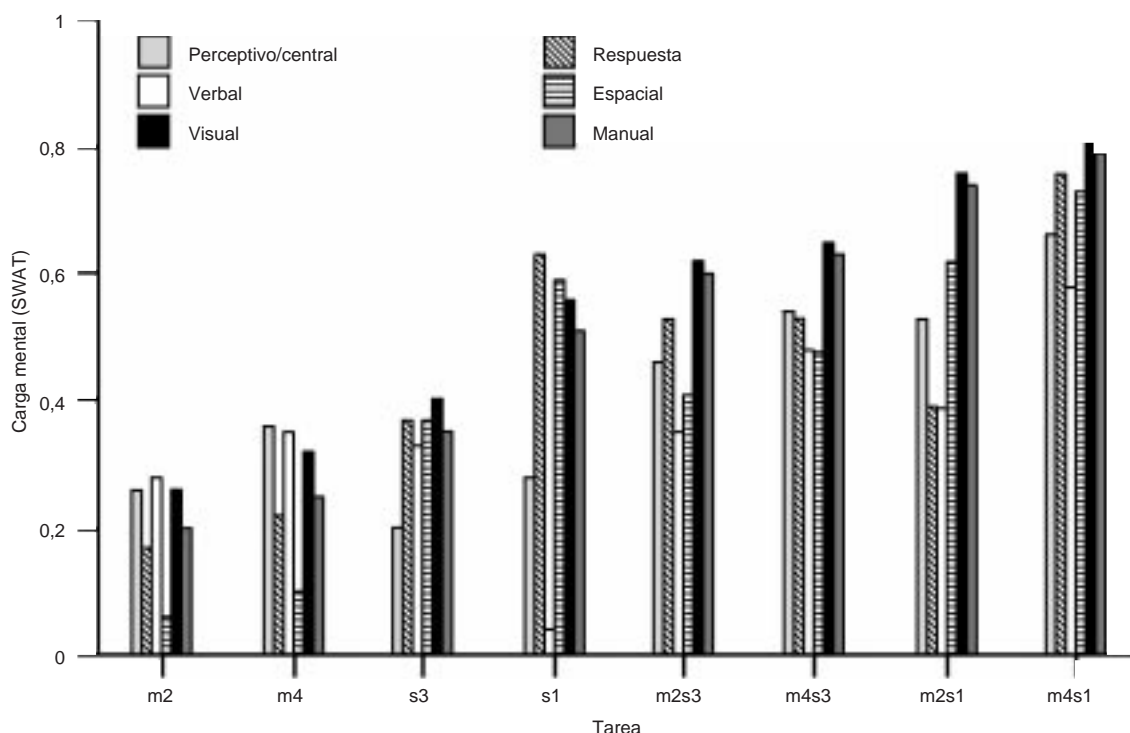


Figura 5. Perfiles de carga mental en WP (Workload Profile).

Siguiendo a Tsang y Velázquez¹⁰, el poder de diagnóstico de cada instrumento fue evaluado en función del grado en el que los perfiles de carga mental servían para discriminar entre las diferentes tareas. Con respecto a los instrumentos NASA-TLX y SWAT, ambos generaron agrupamientos muy similares a las tareas, si bien el poder discriminante de la técnica SWAT es muy superior al del NASA-TLX. En cuanto al WP, podemos concluir que la primera función discriminante obtenida discrimina las tareas simples de memoria de las tareas simples de seguimiento. Tomando como referencia el modelo de recursos múltiples de Wickens⁴, podemos concluir que las coordenadas en el espacio bidimensional de la tarea de seguimiento indican que esta tarea demanda recursos de procesamiento de respuesta y su código de procesamiento es espacial. Sin embargo, la tarea de memoria demanda recursos de procesamiento perceptivo/central y su código de procesamiento es verbal. Además, la segunda función discrimina entre tareas simples y duales. Teniendo en cuenta que las tareas simples obtienen puntuaciones negativas y las duales positivas, podemos concluir que las tareas duales demandan recursos de procesamiento de todos los tipos, mientras que las simples sólo demandan algunos de ellos.

Estos resultados confirman los obtenidos por Tsang y Velázquez¹⁰, es decir, el elevado poder de diagnóstico del instrumento WP, ya que los perfiles de carga mental obtenidos con esta técnica fueron capaces de detectar con gran precisión las diferencias en el tipo de recursos atencionales demandados por cada una de las tareas, tal y como defiende el modelo de recursos múltiples⁴.

Por otro lado, aunque no se han cuantificado las diferencias entre los tres instrumentos en relación a sus requisitos de implementación y su grado de aceptación por los sujetos, en la aplicación de cada uno de ellos se pudieron observar algunas ventajas de unos frente a los otros. En general, al tratarse de técnicas de papel y lápiz, los requisitos de implementación de las tres técnicas son mínimos. La única diferencia apreciable se refiere al tiempo de aplicación. En este sentido, la técnica SWAT requirió un tiempo algo superior (aproximadamente 75 min) a las otras dos (aproximadamente 60 min).

En cuanto a la aceptación por parte de los sujetos, los tres instrumentos fueron bien aceptados por todos los participantes en esta investigación, aunque con el WP los individuos manifestaron algunos problemas en la com-

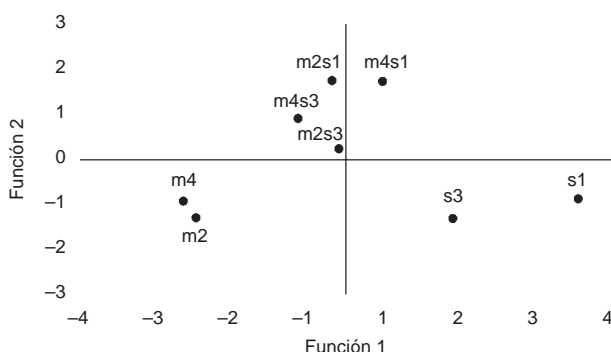


Figura 6. Resultados del análisis discriminante para el instrumento WP.

preensión de las dimensiones que distingue este instrumento. En cuanto al SWAT, la tarea de ordenación previa a la realización de las tareas experimentales resultó algo tediosa para los sujetos.

En resumen, y como conclusiones de carácter más general, se pueden establecer las siguientes recomendaciones básicas en función de los objetivos que se persigan en la evaluación de la carga mental en contextos aplicados. Así, si el objetivo es comparar la carga mental de dos o más tareas que difieren en parámetros objetivos de dificultad, el evaluador podrá optar por cualquiera de las tres técnicas evaluadas, aunque la elección más idónea para este fin es la del Perfil de Carga. A su vez, si lo que se pretende es realizar un análisis de las demandas cognitivas o de los recursos atencionales que demanda una determinada tarea, con el fin de diseñar la distribución óptima de las tareas y la situación de trabajo que produzcan niveles de rendimiento más elevados, la elección más adecuada sería la del WP, seguida del SWAT.

AGRADECIMIENTOS

Este artículo representa una parte de la investigación que obtuvo el segundo «Premio de Investigación Psicológica General González del Pino», otorgado por el Ministerio de Defensa español en 1998.

BIBLIOGRAFÍA

1. Navon D, Gopher D. On the economy of the human processing system. *Psychol Rev* 1979; 86: 214-255
2. O'Donnell R, Eggemeier FT. Workload assessment methodology. En: Boff KR, Kaufman L, Thomas JP, editores. *Handbook of perception and human performance*. Nueva York: Wiley, 1986; 1-49.
3. Gopher D, Donchin E. Workload: an examination of the concept. En: Boff KR, Kaufman L, Thomas JP, editores. *Handbook of perception and human performance*. Nueva York: Wiley & Sons, 1986.
4. Wickens CD. *Engineering psychology and human performance*. Nueva York: Harper Collins, 1992.
5. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. En: Hancock PA, Meshkati N, editores. *Human Mental Workload*. Amsterdam: Elsevier, 1988, 139-183.
6. Reid GB, Eggemeier FT, Shingledecker CA. Subjective workload assessment technique. *Proceedings of the 1982 AIAA Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics*, 1982; 281-288.
7. Eggemeier FT, Wilson GF, Kramer AF, Damos DL. General considerations concerning workload assessment in multi-task environments. En: Damos DL, editor. *Multiple task performance*. Londres: Taylor & Francis, 1991; 207-216.
8. Meshkati N, Hancock P, Rahimi M. Techniques in mental workload assessment. En: *Evaluation of human work*. University of Nottingham, 1992.
9. Reid GB, Nygren TE. The subjective workload assessment technique: a scaling procedure for measuring mental workload. En: Hancock PA, Meshkati N, editores, *Human mental workload*. Amsterdam: Elsevier, 1988; 185-218.
10. Tsang PS, Velázquez VL. Diagnosticity and multidimensional subjective workload ratings. *Ergonomics* 1996; 39: 358-381.
11. Nygren TE. Psychometric properties of subjective workload techniques: implications for their use in the assessment of perceived mental workload. *Human Factors* 1991; 33: 17-33.