

ÍNDICES DE DIFICULTAD SUBJETIVOS Y FISIOLÓGICOS EN LA PRUEBA DE MATRICES DE RAVEN

M. REDONDO; F. DEL VALLE-INCLÁN
Universidad de La Coruña

Resumen

Se estudian las variaciones en medidas de esfuerzo mental subjetivas y fisiológicas durante la resolución de dos series, con diferente dificultad, del test de matrices progresivas de Raven. Como medida subjetiva se utilizó una escala con anclas verbales. Las medidas fisiológicas fueron estimaciones de la variabilidad cardíaca en el dominio del tiempo y de la frecuencia.

Los resultados muestran que, en función de la dificultad, se incrementan el tiempo empleado y el esfuerzo percibido, y decrece el número de respuestas correctas. Las estimaciones de variabilidad cardíaca mostraron un decremento en función de la dificultad, especialmente en la banda de frecuencias medias, que parece ser más sensible a la dificultad de la tarea. Los resultados de variabilidad cardíaca extienden los hallazgos previos de otros autores. Aunque las medidas fisiológicas y subjetivas se mostraron sensibles a la dificultad de la tarea, no se encontraron correlaciones entre ellas, lo cual es un dato ampliamente repetido en la literatura.

Abstract

Subjective and physiological measures of mental effort were taken for two series of Raven test differing in difficulty. The subjective measure was a rating scale with verbal anchors. The physiological measures were estimators of heart-rate variability (HRV) in time and frequency domains.

The results showed an increase in time required to solve the series and in perceived effort, along with a decrement in the number of correct responses, during the difficult task. HRV showed a tendency to decrease with increasing difficulty, but only reliable effects were obtained in the amplitude of the mid-frequency band. These results extend previous work. Although subjective and HRV measures showed sensibility to task difficulty, they did not correlate. This lack of correlation is the usual result of experiments comparing physiological and subjective measures.

Introducción

La evaluación de la demanda atencional, o la dificultad de la tarea, es un aspecto importante del diseño de «interfaces» hombre-máquina, especialmente cuando la interacción es compleja. Medidas para este propósito son usuales, por ejemplo, en la industria aeronáutica, donde se han desarrollado varias escalas para medir las demandas del sistema sobre el piloto (véase las revisiones de O'Donnell y Eggemeier, 1986; Gopher y Donchin, 1986). Para evaluar las demandas atencionales se utilizan medidas subjetivas (escalas, cuestionarios, autoinformes en general), conductuales (velocidad y/o precisión de la respuesta en dobles tareas) y medidas fisiológicas. Los índices fisiológicos que parecen más sensibles son: el diámetro pupilar (véase la revisión de

Beatty, 1982), la amplitud de P300 (Isreal y otros, 1980) y la variabilidad cardíaca (Mulder, 1980; Mulder, 1988). De todos ellos, la variabilidad cardíaca es el menos aparatoso y hay evidencias, provenientes de laboratorios y situaciones de campo, de que se comporta como un índice de dificultad de la tarea.

Se entiende por «variabilidad cardíaca» la variación en el tiempo entre dos contracciones ventriculares. La utilización de esta medida como índice de la dificultad de la tarea se remonta a los trabajos de Kalsbeek y Etema (1963). Estos autores mostraron cómo a medida que aumentaba la carga, los sujetos mostraban una disminución de la variabilidad cardíaca. Kalsbeek y Etema (1963) utilizaron una tarea donde dos estímulos se asociaban con dos respuestas diferentes y variaron el intervalo entre estímulos.

Comprobaron que a medida que aumentaba la frecuencia de estimulación (es decir, la dificultad de la tarea, o la carga mental) disminuía la variabilidad cardíaca.

La variabilidad cardíaca se estima en el dominio del tiempo y de la frecuencia. En el primero, el índice que parece tener más aceptación es el coeficiente de variación, mientras que los análisis en el dominio de la frecuencia han mostrado la existencia de tres bandas:

1. La banda de bajas frecuencias (0,01-0,06 Hz) relacionada con mecanismos de vasodilatación periférica, y por tanto con la regulación de la temperatura corporal (Sayers, 1980).

2. La banda de frecuencias medias (0,07-0,14 Hz), o banda de presión arterial, refleja la actividad de los mecanismos barorreceptores y su influencia sobre el corazón a través del sistema nervioso parasimpático. Mulder y Mulder (1981a, 1981b) han mostrado cómo la banda media es más específica, en lo que a esfuerzo mental se refiere, que el resto del espectro.

3. Por último, la banda de frecuencias respiratorias con una frecuencia, generalmente, y en condiciones de laboratorio psicológico, superior a 0,20 Hz.

Las medidas subjetivas tienen obvios inconvenientes, como son las creencias del sujeto sobre el experimento, o lo que se espera de él, sus motivaciones, etc.; en suma, factores que no controla ni conoce el experimentador. Pese a ello, hay autores que defienden que el autoinforme es, en realidad, el criterio último. Si las personas dicen que están, o que no están, sobrecargadas por una tarea, los resultados con cualquier otro índice deberán ser coincidentes, y si no lo son habrán de considerarse insensibles (Moray, 1979). Para la medida del esfuerzo percibido se han construido muchos instrumentos: cuestionarios, escalas, mezclas de ambos, pero no parece haber grandes diferencias entre unos u otros métodos (Borg, 1978).

El experimento que presentamos se diseñó para estudiar las variaciones en medidas subjetivas y fisiológicas (variabilidad cardíaca) durante la resolución del test de matrices progresivas de Raven. Las matrices de Raven se han utilizado en experimentos de estimación del esfuerzo mental por Borg (1978) y Borg, Bratfish y Dornic (1971), entre otros. Prácticamente todos los trabajos provienen de la Universidad de Estocolmo y muestran que la prueba de Raven es un buen instrumento para producir cargas mentales diferentes (medidas por autoinforme).

Método

Sujetos

Un total de 21 alumnos de Psicología participaron voluntariamente en el experimento. La media de

edad fue de 21,5 años (SD = 3,5) y el 70 por 100 eran mujeres. Ninguno sufría enfermedades cardiovasculares ni tomaba medicación.

Estímulos y medidas subjetivas

Los estímulos se extrajeron de las series A, B, C y E de la prueba de Raven para adultos (forma general). Había cuatro bloques de doce estímulos cada uno: los dos primeros eran una mezcla al 50 por 100 de las series A y B (pares de A e impares de B en uno, y pares de B e impares de A, en otro), y por tanto no diferían en dificultad; los dos segundos correspondían a las series C y E, sin modificar. La medida del esfuerzo subjetivo se realizó mediante una escala con anclas verbales elaborada por Redondo (1990) siguiendo a Zijlstra (1988).

Registro de la variabilidad cardíaca

El ECG se registró en una derivación precordial. La señal amplificada pasaba por una *trigger* que daba un pulso en cada onda R (el signo eléctrico de la contracción ventricular). El momento en que se producía un pulso se registraba con precisión de 1 milisegundo en un *event-data collector* digital. La secuencia de eventos se pasaba después a un PC para su almacenamiento y análisis.

Procedimiento

El experimento se realizó individualmente. Los sujetos se sentaban en una silla y se les ponían los electrodos. Delante de ellos, sobre una mesa, colocados boca abajo, estaban las hojas con las matrices. Se instruyó a los sujetos para que levantasen las hojas de una en una, anotasen la solución en una hoja de respuestas y pasasen a la siguiente, en el menor tiempo posible. Al término de cada bloque, evaluaban el esfuerzo percibido. El experimentador permanecía en la sala y medía el tiempo empleado en resolver cada bloque. Los dos primeros bloques eran de una dificultad similar y se consideraron como práctica. Los otros dos bloques diferían en dificultad y se presentaron contrabalanceados.

Análisis de datos

Para cada bloque y sujeto, se analizó el tiempo invertido, el número de soluciones correctas, las puntuaciones de esfuerzo dadas y las estimaciones de variabilidad cardíaca. El análisis de la variabilidad cardíaca se realizó con CARSPAN (Meulen y Mulder, 1988). La unidad de medida en el dominio del tiempo fue el IBI (*inter-beat interval*) en milisegundos, y la estimación de variabilidad el coeficiente de variación. Los análisis espectrales se realizaron con la frecuencia cardíaca y se normalizó la amplitud (véase Meulen y Mulder, 1988). Además de la ener-

gía espectral total (0-0,50 Hz), se calculó la amplitud en las bandas de frecuencias bajas (0,01-0,06 Hz) y medias (0,07-0,14 Hz). Las estimaciones espectrales se transformaron en su logaritmo natural antes de ser analizadas estadísticamente.

Se realizó un ANOVA para cada variable, siendo los factores: orden de presentación (2), sexo (2) y dificultad de la tarea (2). La variable «sexo» se introdujo en el análisis porque se sabe que los hombres tienen una variabilidad mayor que las mujeres (véase, por ejemplo, Mulder, 1980).

Resultados

Los resultados conductuales y subjetivos se exponen en la tabla 1. Como se aprecia, el esfuerzo percibido y el tiempo empleado en resolver la tarea se incrementaron con la dificultad, mientras el número de problemas solucionados correctamente decreció. El análisis estadístico mostró un efecto significativo de la dificultad de la tarea en las tres variables: estimaciones de esfuerzo ($F(1,17) = 45,63$, $p < 0,0001$), rendimiento ($F(1,17) = 7,30$, $p < 0,01$) y tiempo empleado en resolver un bloque de doce problemas ($F(1,17) = 58,03$, $p < 0,0001$).

TABLA 1

Resultados de autoevaluaciones de esfuerzo, número de respuestas correctas (rendimiento) y tiempo empleado (en segundos)

Tipo de tarea	Esfuerzo	Rendimiento	Tiempo
Fácil	43,762	9,333	302,524
Difícil	67,524	7,571	518,190

Respecto a las estimaciones de variabilidad cardíaca, como era de esperar, fueron más altas para el grupo de hombres. El IBI fue más alto en los hombres que en las mujeres ($F(1,17) = 5,75$, $p < 0,03$). Las medidas de densidad espectral también tuvieron valores más altos en hombres que en mujeres para el espectro total ($F(1,17) = 4,37$, $p < 0,05$) y para la banda de frecuencias medias ($F(1,17) = 4,25$, $p < 0,05$). La figura 1 muestra la amplitud de la banda media a lo largo de la sesión para los dos órdenes experimentales. Se aprecia que uno de los grupos tiene valores más altos, lo que se debe a que hay más hombres en ese grupo. Se aprecia también que las condiciones difíciles tienen menor amplitud que las fáciles, independientemente del orden de presentación.

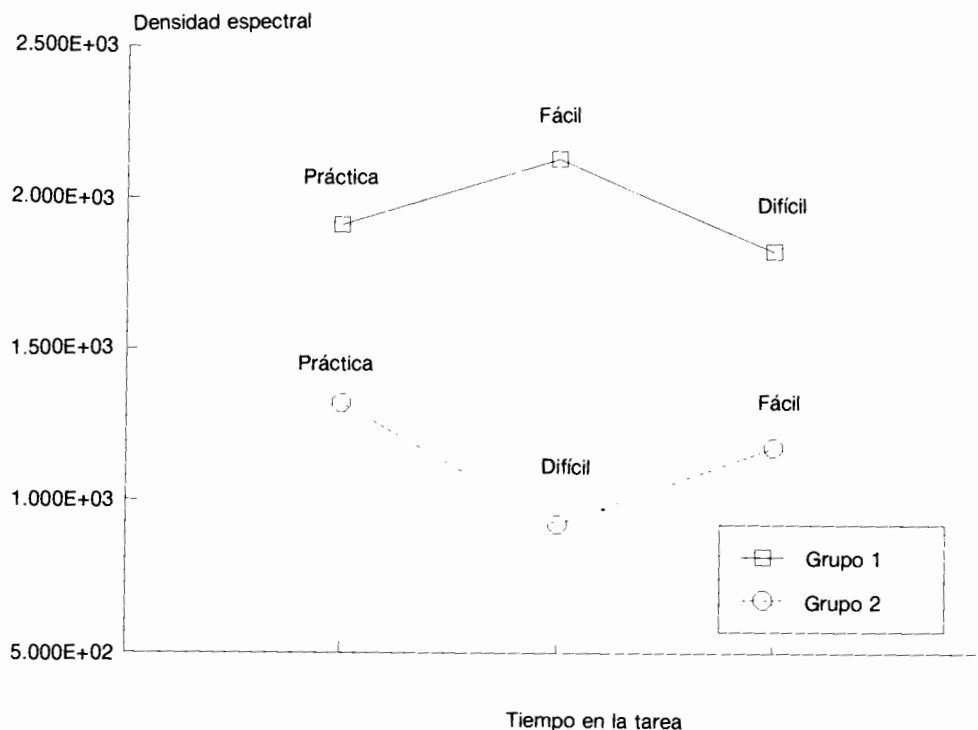


Figura 1. Cambios de amplitud en la banda de frecuencia medias a lo largo de la tarea, para los dos órdenes de presentación. Los valores de «práctica» son el promedio de los dos períodos de práctica.

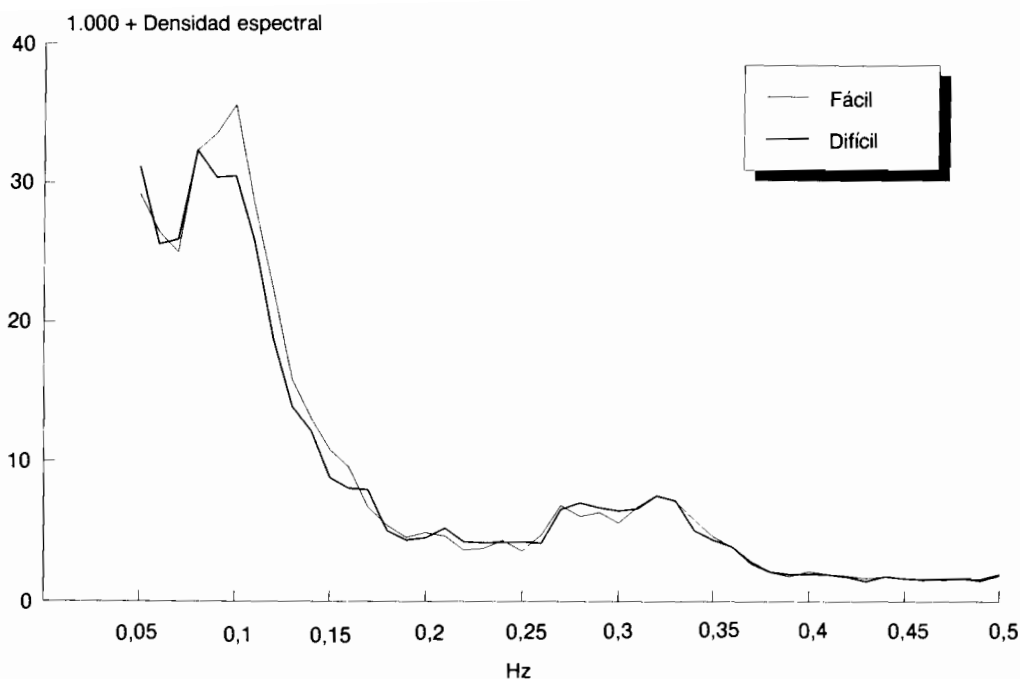


Figura 2. Densidad espectral (0,04 – 0,50 Hz) para las dos condiciones experimentales.

La figura 2 recoge las funciones de densidad espectral entre 0,04 y 0,50 Hz. Se observan dos picos que corresponden a la banda de frecuencias medias (alrededor de 0,1 Hz) y a la frecuencia respiratoria (0,3 Hz), y la única diferencia entre las dos condiciones experimentales parece estar en la menor amplitud del componente de 0,1 Hz en la tarea difícil. El ANOVA mostró que esta observación era correcta ($F(1,17) = 4,80, p < 0,04$). Se encontró, además, un decremento de la densidad espectral total ($F(1,17) = 5,05, p < 0,04$) en función de la dificultad.

Como análisis adicional, y con el objetivo de comparar las diferentes medidas entre sí, realizamos para cada variable las diferencias entre las condiciones fácil y difícil, y con ellas calculamos las correlaciones de Pearson entre las medidas. Las estimaciones subjetivas correlacionaron positivamente con el tiempo empleado ($r = 0,5426, p < 0,01$) y negativamente con el rendimiento ($r = -0,607, p < 0,01$). Como era de esperar, los índices de variabilidad cardíaca correlacionaron entre sí positivamente, pero no hubo correlaciones con las medidas conductuales o subjetivas.

Discusión

Las medidas conductuales reflejaron la dificultad de la tarea: el tiempo empleado en resolver los doce

problemas se incrementó con la dificultad, y el número de respuestas correctas decreció. La medida subjetiva de esfuerzo también aumentó con la dificultad y, además, mostró correlaciones con las dos variables conductuales. La correlación positiva entre tiempo empleado y esfuerzo y la correlación negativa entre esfuerzo y rendimiento son las que cabe esperar, y muestran que la escala parece ser un buen instrumento para evaluar el esfuerzo de los sujetos.

Los resultados de la variabilidad cardíaca vienen a confirmar y extender los resultados de Mulder y Mulder (1981a, 1982b) y Aasman, Mulder y Mulder (1987), entre otros, que utilizaron tareas QRST (Massaro, 1975) y encontraron decrementos en la banda de frecuencias medias y en la energía espectral total al aumentar la dificultad de la tarea. Puesto que prácticamente todos los estudios de laboratorio sobre las relaciones entre variabilidad cardíaca y dificultad de la tarea provienen de tareas de tiempos de reacción, la obtención de resultados similares en contextos diferentes, como el de este experimento, viene a confirmar la relación entre dificultad de la tarea y decrementos de variabilidad cardíaca, especialmente en la amplitud de la banda de frecuencias medias (alrededor de 0,1 Hz).

La ausencia de correlaciones entre las medidas fisiológicas y las subjetivas es un hallazgo común en la literatura e independiente de las formas de medida. Por ejemplo, Casali y Wierwille (1983, 1984) y Hicks y Wierwille (1979) encontraron correlaciones

muy bajas. Moray (1982) señala que es relativamente raro encontrar correlaciones entre estos tipos de medidas. Si bien es un hallazgo usual, su interpretación es problemática. Aunque el valor de las medidas de autoinforme sea cuestionable desde puntos de vista diferentes, el argumento de Moray (1979) es difícilmente controvertible: si una persona dice que algo le cuesta esfuerzo, es irrelevante que las medidas conductuales y fisiológicas no lo reflejen (un argumento similar ha empleado Navon (1986), al discutir las medidas más adecuadas para la percepción consciente).

Nota: Este trabajo fue posible gracias a la financiación de la Consellería de Educación (Xunta de Galicia) y a la amabilidad del profesor G. Mulder, de la Universidad de Groningen (Holanda).

Referencias

- Aasman, J., Mulder, G. y Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors*, 29, 161-170.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load and the structure of processing resources. *Psychological Bulletin*, 91, 276-292.
- Borg, G. (1978). Subjective aspects of physical and mental load. *Ergonomics*, 21, 215-220.
- Borg, G., Bratfisch, O. y Dornic, S. (1971). On the problems of perceived difficulty. *Scandinavian Journal of Psychology*, 12, 249-260.
- Casali, J. C. y Wierwille, W. W. (1983). A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors*, 25, 623-641.
- Casali, J. C. y Wierwille, W. W. (1984). On the measurement of pilot perceptual workload: A comparison of assessment techniques addressing sensitivity and intrusion issues. *Ergonomics*, 27, 1033-1055.
- Gopher, D. y Donchin, E. (1986). Workload: An examination of the concept. En K. R. Boff, L. Kaufman y J. P. Thomas (Eds.), *Handbook of Perception and Human Performance*, vol. II: *Cognitive Processes and Performance*, cap. 41. New York: Wiley.
- Hicks, T. y Wierwille, W. (1979). Comparison of five mental workload assessment procedures in a moving-base simulator. *Human Factors*, 21, 129-144.
- Isreal, J. B., Wickens, C. D., Chesney, G. L. y Donchin, E. (1980). The event-related brain potential as an index of display-monitoring workload. *Human Factors*, 22, 212-224.
- Kalsbeek, J. W. H. y Etema, J. H. (1963). Continuous recording of heart rate and the measurement of perceptual load. *Ergonomics*, 6, 306-307.
- Massaro, D. W. (1975). *Experimental Psychology and Information Processing*. Chicago: Rand McNally.
- Meulen, P. y Mulder, L. J. M. (1987). CARSPAN, Institute for Experimental and Occupational Psychology, University of Groningen, Holanda.
- Moray, N. (1979). Models and measures of mental workload. En Moray, N. (Ed.), *Mental Workload: Its Theory and Measurement*. New York: Plenum Press.
- Moray, N. (1982). Subjective mental workload. *Human Factors*, 24, 25-40.
- Mulder, G. (1980). *The Heart of Mental Effort: Studies in the Cardiovascular Psychophysiology of Mental Work*. Tesis doctoral. Universidad de Groningen, Holanda.
- Mulder, G. y Mulder, L. J. M. (1981a). Information processing and cardiovascular control. *Psychophysiology*, 18, 392-405.
- Mulder, G. y Mulder, L. J. M. (1981b). Task-related cardiovascular stress. En J. Long y A. D. Baddeley (Eds.), *Attention and Performance*, vol. 9, Hillsdale, N.J.: Erlbaum.
- Mulder, L. J. M. (1988). *Assessment of cardiovascular reactivity by means of spectral analysis*. Tesis doctoral. Universidad de Groningen, Holanda.
- Navon, D. (1986). On determining what is unconscious and what is perception. *Behavioral and Brain Sciences*, 9, 44-45.
- O'Donnell, R. D. y Eggemeier, E. T. (1986). Workload assessment methodology. En K. R. Boff, L. Kaufman y J. P. Thomas (Eds.), *Handbook of Perception and Human Performance*, vol. 2. N. Y.: Wiley & Sons.
- Redondo, M. (1990). La construcción de una escala para medir esfuerzo mental. Datos no publicados.
- Sayers, B. McA. (1980). Pattern analysis of the heart rate signal. En I. Martin y P. H. Venables (Eds.), *Techniques in Psychophysiology*, 197-210. N. Y.: Wiley & Sons.
- Zijlstra, F. (1988). Mental-effort scale. Informe interno. Institute for Experimental and Occupational Psychology. Universidad de Groningen. Groningen, Holanda.