Psicothema

Psicothema 2017, Vol. 29, No. 4, 570-576
doi: 10.7334/psicothema2017.151

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Development of the CarMen-Q Questionnaire for mental workload assessment

Susana Rubio-Valdehita, María Inmaculada López-Núñez, Ramón López-Higes and Eva María Díaz-Ramiro
Universidad Complutense de Madrid

Abstract

Background: Mental workload has emerged as one of the most important occupational risk factors present in most psychological and physical diseases caused by work. In view of the lack of specific tools to assess mental workload, the objective of this research was to assess the construct validity and reliability of a new questionnaire for mental workload assessment (CarMen-Q). **Method:** The sample was composed of 884 workers from several professional sectors, between 18 and 65 years old, 53.4% men and 46.6% women. To evaluate the validity based on relationships with other measures, the NASA-TLX scale was also administered. **Results:** Confirmatory factor analysis showed an internal structure made up of four dimensions: cognitive, temporal and emotional demands and performance requirement. The results show satisfactory evidence of validity based on relationships with NASA-TLX and good reliability. **Conclusions:** The questionnaire has good psychometric properties and can be an easy, brief, useful tool for mental workload diagnosis and prevention.

Keywords: Mental Workload, occupational ergonomics, questionnaire, validation.

Resumen

Desarrollo del cuestionario CarMen-Q para evaluar la carga mental de trabajo. **Antecedentes:** actualmente la carga mental ha surgido como uno de los factores de riesgo laboral más importantes presentes en la mayoría de las enfermedades psicológicas y físicas causadas por el trabajo. Ante la falta de herramientas específicas para evaluar la carga mental, el objetivo de esta investigación fue evaluar la validez de constructo y la fiabilidad de un nuevo cuestionario (CarMen-Q) para la evaluación de la carga mental de trabajo. **Método:** la muestra estuvo formada por 884 trabajadores de diversos sectores profesionales, de entre 18 y 65 años de edad, de los cuales el 53,4% fueron hombres y el 46,6% mujeres. Para evaluar la validez basada en las relaciones con otras medidas también se administró la escala NASA-TLX. **Resultados:** el análisis factorial confirmatorio mostró una estructura interna formada por cuatro dimensiones: demandas cognitiva, temporal y emocional y requisitos de rendimiento. Se encontró evidencia satisfactoria de validez basada en las relaciones con la escala NASA-TLX y adecuados índices de fiabilidad. **Conclusiones:** el cuestionario tiene buenas propiedades psicométricas y es una herramienta sencilla y breve, útil para el diagnóstico y la prevención de la carga mental.

Palabras clave: carga mental, ergonomía laboral, cuestionario, validación.

Psychosocial risk factors are conditions in a work context which can affect workers' welfare, health and efficiency. These factors include mental workload (MW) as one of the most important. Exposure to MW can lead to serious health problems for workers (cardiovascular diseases, digestive problems, anxiety, burnout, etc.). In a meta-analytic review, Stansfeld and Candy (2006) showed the relationships between MW (and other psychosocial work environment factors) and mental health. The most direct consequence of MW on workers' health and well-being is occupational stress (Goh, Pfeffer, & Zenios, 2015).

MW is one of the most widely used concepts in ergonomics and human factors and represents a topic of increasing importance. Since modern technology in many working environments imposes ever more cognitive demands while physical demands diminish, understanding how MW affects performance and psychological workers' well-being is increasingly critical (Young, Brookhuis, Wickens, & Hancock, 2015). Excess workload can result in slower task performance and errors and it should also be noted that underload can also lead to boredom, loss of situation awareness and reduced alertness. Workload may be more relevant in times of economising or temporarily during peaks (such as incidents or turnarounds). A perceived high workload not only adversely affects safety, but also negatively disturbs job satisfaction and, as a result, contributes to high turnover and staff shortages.

There is no clearly defined, universally accepted definition of MW. Aspects of MW seem to fall within three broad categories: the amount of work and number of tasks to do, the particular aspect of time one is concerned with and the subjective psychological experiences of the human. MW is a construct reflecting the interaction of demands imposed on workers by tasks they have to perform and cannot be directly observed, it must be inferred from observation of overt behaviour or measurement of psychological and physiological processes. Workers' capabilities and effort in the

Received: April 20, 2017 • Accepted: June 29, 2017
Corresponding author: Susana Rubio-Valdehita
Facultad de Psicología
Universidad Complutense de Madrid
28223 Madrid (Spain)
e-mail: srubiova@ucm.es

context of specific situations all moderate the MW experienced by the individual. MW is multidimensional and multifaceted and results from the aggregation of many different demands and so is difficult to define uniquely (Cain, 2007). Verwey (2000) points out that MW experienced depends on individual abilities, motivation, strategies applied to perform a task and physical and emotional state. Hart (2006) states that MW is result of the interaction between task requirements, the circumstances in which it is carried out, and subject's abilities, emotions and perceptions. In order to understand more completely the MW construct, the model proposed by González (2003) includes the emotional aspects involved in this process. As a consequence of the person-job interaction, there may be a positive level of stress (eustres) or negative (distres) related to excessive activation and an overflowing perception of demands. In this way, when a high activation occurs, the person's information processing capacity decreases and his MW increases as long as the person keeps trying to reach the same performance.

Different methods have been proposed to assess MW: performance indicators, physiological measures, scales and questionnaires. Most of theoretical models of MW highlight the need to focused on subjective aspects, since worker's subjective perception will determine his/her satisfaction and well-being (Szalma, 2008). In addition, physiological and performance techniques involve the use of intrusive methods that obstruct reliable MW assessment in real work situations, becoming rejected by workers.

NASA-TLX scale (Hart & Staveland, 1988) is the most widely used instrument to measure MW. It distinguishes six dimensions on MW (effort, mental demand, physical demand, temporal demand, performance and frustration) that are evaluated from 0 to 100. NASA-TLX has proven to be a sensitive, valid and reliable tool (Rubio, Díaz, Martín, & Puente, 2004). However, TLX has practical limitations, especially in relation to its presentation format, which is unusual in real work areas. Workers are much more familiarized with questionnaires consisting of a series of items which clearly ask them to express their opinion about the intensity, importance or frequency with which a particular condition referred to in a specific item is presented at workplace, using a Likert scale. In Spain there is no specific instrument to measure MW with appropriate psychometric guarantees and in a simple and easy way. Frequently, some measures associated to MW are introduced as a part of a general method designed to assess several psychosocial risks as DECORE, ISTAS-21 or F-PSICO, forcing occupational health professionals to perform complex and long-time consuming assessments and in most cases their psychometric properties have not been sufficiently tested (Moreno & Báez, 2010). In 2009 Rolo-González, Díaz-Cabrera, & Hernández-Fernaud, proposed ESCAM Scale in an attempt to obtain scores on several dimensions of workload, but as their authors pointed out, its utility is still limited and needs further development.

The main objective of this paper is to present a psychometric study focused in a new tool for MW evaluation and diagnosis, that we are denominated CarMen-Q (from the first syllabus in spanish of Carga Mental Questionnaire), which pretends to be a valid and reliable questionnaire, easy to apply in real-work contexts covering the lack of specific MW assessment tools in Spain. In CarMen-Q's development we have followed some of the assumptions of Hart and Staveland's (1988) MW model, thus assuming construct's multidimensional nature and a structure formed by aspects associated to task's features (mental, physical and temporal demands), performance requirements and emotional impact of the task demands on the worker (frustration) (Chiorri, Garbarino, Bracco, & Magnavita, 2015). Considering previous research on NASA-TLX (DiDomenico & Nussbaum, 2011) showing a limited practical relevance of its performance dimension (since it is influenced by variations in physical load), CarMen-Q revises that dimension and does not include items associated to job's physical aspects in order to obtain a pure measure of MW. Furthermore, performance may not always deteriorate with an increase in MW simply because the individual, as a result of evaluating the difference between demand and performance, invests more resources in order to satisfy task demands (Yeh & Wickens, 1988). Subjective assessments of MW may not provide accurate estimation of performance. Instead, subjective measures may be more useful in indicating potential performance problems if task demands are further increased (Yeh & Wickens, 1988), so CarMen-Q is designed to measure job's performance requirement rather than worker's performance level.

Method

Participants

The study involved an incidental sample of 884 workers from different professional sectors. Participation in the study was offered to 1041 workers, of whom 957 agreed to participate (response rate = 91.93%). Of these, 73 were eliminated because presented some missing response in some instrument. Mean age of final sample was 41.21 years (SD = 10.45), ranging from 18 to 65 years old, 53.4% males and 46.6% females. Regarding civil status, 29.9% are single, 63.8% are married or have a stable partner, and 6.3% are separated, divorced or widowed; 57.7% of participants have between one and 6 children, being two the most frequent number of children (53.3%). The average antiqueness in job was 12.66 years (*SD* = 9.43). To participate in the study it was necessary to have an antiqueness in job of at least six months. With respect to professional sector, 46.1% are health professionals, 24.7% administrative staff, 16.3% security personnel and 12.9% works in service sector.

Instruments

CarMen-Q was developed to distinguish four MW dimensions: cognitive, temporal, and emotional/health and performance demands. Performance dimension was formulated asking workers about their job's performance requirements instead of about the estimated performance level that he/she reaches (as NASA-TLX does). CarMen-Q does not includes items asking for physical demands.

CarMen-Q is a paper-and-pencil questionnaire designed to assess MW in a simple, valid and reliable way and consists of 29 items (Table 1). Item's response format is a Likert frequency scale of four alternatives in which 0 means never, 1 rarely, 2 often and 3 always. All items have been stated so that a higher score indicates a higher mental load. Steps proposed by Downing (2006) were followed in CarMen-Q development. In this sense, at first a panel of seven experts in occupational ergonomics and MW met to elaborate a series of items that evaluated the following dimensions: cognitive and temporal demands, performance requirements and emotional aspects. As a result a questionnaire of 66 items was applied to a pilot sample of 300 workers. After factor and reliability analysis, 16 items were eliminated as they showed inadequate adjustment indexes. The second version was applied to a sample of 200 workers, and the results of reliability and validity showed the convenience of eliminating again a series of items, giving rise to the final version of 29 items that is used in this study.

To analyze validity of CarMen-Q based on relationships with other measures we also applied NASA-TLX (Hart & Staveland, 1988), in its Spanish adaptation (Rubio et al., 2004). NASA-TLX is a multidimensional instrument that distinguish six dimensions evaluated on a 0-100 scale: mental, physical and temporal demands, performance level, effort, and frustration. A number between 0 and 100 is selected in a graphical scale according to the perceived workload in each subscale, and all the scores are averaged to determine total workload. NASA-TLX is the most widely used instruments to assess subjective workload. NASA-TLX is popular because it is simpler, has adequate psychometric properties (Cronbach's alpha around .70) and high sensibility (Rubio et al., 2004). In our sample, NASA-TLX showed a Cronbach's alpha equal to .78.

Procedure

Before data collection, authors contacted to the responsible for the occupational risk prevention services of participant's work centres. Once instrument application was authorized, participants were informed about objectives of the study and signed an informed consent document, explaining the anonymous and voluntary nature of their participation. The application session was organized by work-centres' managers in collaboration with the authors, it was collective and each session lasted about 15 minutes. Ethics committee of the authors' research centre approved the study.

Data analysis

To evaluate validity evidence based on internal structure of CarMen-Q, total sample was randomly divided in two subsamples ($n_1 = 317$; $n_2 = 567$), maintaining in both the same mean sociographic characteristics and the percentage of representation of the different professional sectors presented in the total sample. Exploratory factor analysis (EFA) with Oblimin rotation was performed with the first subsample and the resulted structure was cross-validated with the second subsample using confirmatory factor analysis (CFA) with maximum likelihood estimation method. Data's suitability for structure detection was tested by KMO (Kaiser-Meyer-Olkin) measure and Bartlett's test of sphericity. High values of KMO (close to 1.0 and > .50) indicate that a factor analysis may be useful with the data. Small p values (less than 0.05) for Bartlett's test indicate that a factor analysis may be useful with this data. We use the Mardia coefficient considering multivariate normality when its critical ratio is equal or minor than 1.96 (Bian, 2011). As goodness of fit indices, we examined the magnitude of χ^2 divided by its

Table 1
CarMen-Q items (In Spanish)

1. My job requires maintaining a high level of attention *(Mi trabajo requiere mantener un elevado nivel de atención)*
2. My work involves the processing of complex information *(Mi trabajo implica el tratamiento de información compleja)*
3. My job requires thinking and choosing between different alternatives *(Mi trabajo requiere pensar y elegir entre diferentes alternativas)*
4. I have to make difficult decisions *(Tengo que tomar decisiones difíciles)*
5. My job requires handling a lot of knowledge*(Mi trabajo requiere manejar muchos conocimientos)*
6. I have to work constantly, I cannot take breaks beyond strict regulations *(Tengo que trabajar constantemente, no puedo hacer pausas, más allá de las estrictamente reglamentarias)*
7. The pace of work is excessive, difficult to reach even by an experienced worker*(El ritmo de trabajo es excesivo, difícil de alcanzar incluso por un trabajador experimentado)*
8. I often work with annoying interruptions*(Suelo trabajar con interrupciones molestas)*
9. I cannot stop my work when I need it *(No puedo parar o detener mi trabajo cuando lo necesito)*
10. The pace of work is imposed on me *(El ritmo de trabajo me viene impuesto)*
11. The accomplishment of my tasks demands a lot of speed *(La realización de mis tareas exige mucha rapidez)*
12. It is normal for me to accumulate the tasks *(Es normal que se me acumulen las tareas)*
13. My job requires no mistakes *(Mi trabajo requiere que no se cometa ningún error)*
14. I have to give very precise responses *(Tengo que dar respuestas muy precisas)*
15. My mistakes can have serious consequences *(Mis errores pueden tener consecuencias graves)*
16. My job requires dealing with information that is perceived with *difficulty (Mi trabajo requiere tratar con información que se percibe con dificultad)*
17. I have trouble forgetting the problems of my job *(Me cuesta olvidar los problemas de mi trabajo)*
18. My work makes me nervous *(Mi trabajo me pone nervioso)*
19. My work is affecting my personal relationships (family, friends...) *(Mi trabajo está afectando a mis relaciones personales (familia, amigos..))*
20. My job involves a lot of responsibility *(Mi trabajo implica mucha responsabilidad)*
21. I feel very tired, physically fatigued *(Me siento muy cansado, fatigado físicamente)*
22. I have to deal with information that is not easily understood *(Tengo que tratar con información que no se entiende fácilmente)*
23. My job requires a lot of information *(Mi trabajo requiere el tratamiento de gran cantidad de información)*
24. My work affects me a lot emotionally *(Mi trabajo me afecta mucho emocionalmente)*
25. My job requires memorizing a high amount of data *(Mi trabajo requiere memorizar una cantidad elevada de datos)*
26. My work is mentally intense *(Mi trabajo es mentalmente intenso)*
27. I have to do a great search and information gathering to carry out my tasks *(He de realizar una gran búsqueda y recopilación de información para llevar a cabo mis tareas)*
28. When I finish my workday I feel a lot of physical exhaustion *(Al terminar mi jornada laboral siento mucho agotamiento físico)*
29. My work is affecting my health *(Mi trabajo está afectando a mi salud)*

degrees of freedom (CMIN/DF < 3); Root Mean Square Error of Approximation (RMSEA < .05); Standardized Root Mean Residual (SRMR < .08); Corrected goodness index (AGFI), Goodness of Fit (GFI), Normed Fit Index (NFI), Relative Fix Index (RFI) and Parsimony ratio (PRATIO). The values of these indices should be close to .90 or above to be considered a good fit (Tabachnick & Fidell, 2013). There were no missing cases in the sample so it was not necessary to use any imputation method.

Evidences of validity based on relationships with other measures was analyzed by Pearson correlation between NASA-TLX dimensions and the scores obtained as the sum of the items of each CarMen-Q dimensions.

Reliability was analyzed as internal consistency. We considered that reliability indices above .70 are adequate (Nunnally & Bernstein, 1994). Average Variance Extracted (*AVE*) and Composite Reliability (*CR*) were also calculated. A *CR* of .70 or above and an *AVE* of more than .50 are deemed acceptable (Fornell & Larcker, 1981).

All statistical analysis was carried out with SPSS 22.0. Confirmatory factor analysis was performed using AMOS 22.0.

Results

Evidence of validity based on internal structure

Exploratory factor analysis (EFA)

KMO value was .92 and Bartlett's test showed a $p < .001$. Four components with eigenvalues greater than one were obtained (Table 2), explaining 51.91 % of total variance (27.20% the first, 13.18% the second, 6.45% the third and 5.08 % the fourth component). The factor "cognitive demands" consists of 10 items related to the processing of complex information, difficulty in perceiving information, complex decision making, memory load, and the amount of information that needs to be taken into account to perform job tasks. The factor "temporal demands" is constituted by 7 items that ask for work rhythm, presence of annoying interruptions or the possibility of taking breaks when the worker needs it. The factor "emotional demands" is formed by 7 items related to the emotional and health consequences of the job. The "performance demands" factor is formed by 5 items and asks about the performance requirements as level of responsibility, required accuracy of responses and errors severity.

Confirmatory factor analysis (CFA)

The four-factor model obtained after EFA was tested by CFA using Maximum Likelihood estimation method. Critical ratio for Mardia coefficient was 1.47 showing the multivariate normality of the data. Through CFA we tested a model of four first-order factors with these all loading onto one higher-order factor (MW) (Figure 1). Following recommendations of ergonomics experts based on the similar content of the items, covariance between eight pair error measures was liberated. This model showed adequate goodness-of-fit indices: CMIN/DF = 2.16, RMSEA = .040 (LO90 = .031, HI90 = .050); SRMR = .06; AGFI = .95, GFI = .96; NFI = .95; RFI = .94; PRATIO = .90.

Evidences of validity based on relationships with NASA-TLX

CarMen-Q scores were calculated as the sum of all items belonging to each factor. The following Pearson correlation coefficients between NASA-TLX and CarMen-Q scores were significant ($p < .001$): NASA-Mental and CarMen-Cognitive ($r = .73$), NASA-Temporal and CarMen-Temporal ($r = .62$), NASA-frustration and CarMen-emotional ($r = .77$) and NASA-Mental and CarMen performance demands ($r = .59$). Correlations between NASA-TLX physical demand and all CarMen dimensions were non-significant and near to 0. Correlation between NASA-TLX performance dimension and CarMen performance demands was non-significant ($r = .27$).

Reliability

Cronbach's Alpha was near or higher than .80 for the whole scale and subscales. It was stable or decreased if an item was deleted (Table 3).

Composite Reliability (*CR*) for each CarMen-Q dimensions was adequate in all cases (*CR* = .95, cognitive demands; *CR* = .93, emotional demands; *CR* = .90, temporal demands; *CR* = .93, performance demands). Average Variance Extracted (*AVE*) was higher than .50 in all cases, showing the good discriminant validity (*AVE* = .64, cognitive demands; *AVE* = .67, emotional demands; *AVE* = .58, temporal demands; *AVE* = .73, performance demands).

Table 2
EFA results. Structure Matrix

	Component			
	1	2	3	4
Item 23	.788	.201	.295	.255
Item 5	.739	.106	.207	.354
Item 2	.738	.157	.278	.323
Item 27	.732	.251	.164	.165
Item 16	.724	.303	.342	.290
Item 3	.722	.163	.141	.315
Item 4	.711	.238	.177	.330
Item 25	.681	.289	.290	.286
Item 26	.679	.327	.391	.331
Item 22	.653	.322	.328	.141
Item 24	.316	.801	.283	.049
Item 29	.204	.786	.359	.021
Item 19	.230	.745	.301	.033
Item 18	.286	.741	.332	.080
Item 21	.112	.738	.370	.098
Item 17	.354	.711	.311	.137
Item 28	.110	.708	.356	.129
Item 6	.198	.294	.732	.252
Item 9	.204	.387	.729	.139
Item 7	.293	.500	.725	.162
Item 11	.308	.273	.720	.314
Item 10	.181	.358	.670	.230
Item 8	.328	.384	.657	.001
Item 12	.321	.395	.652	.006
Item 13	.287	.103	.275	.814
Item 20	.335	.161	.184	.763
Item 15	.384	.209	.189	.758
Item 14	.300	.132	.243	.725
Item 1	.207	.147	.303	.644

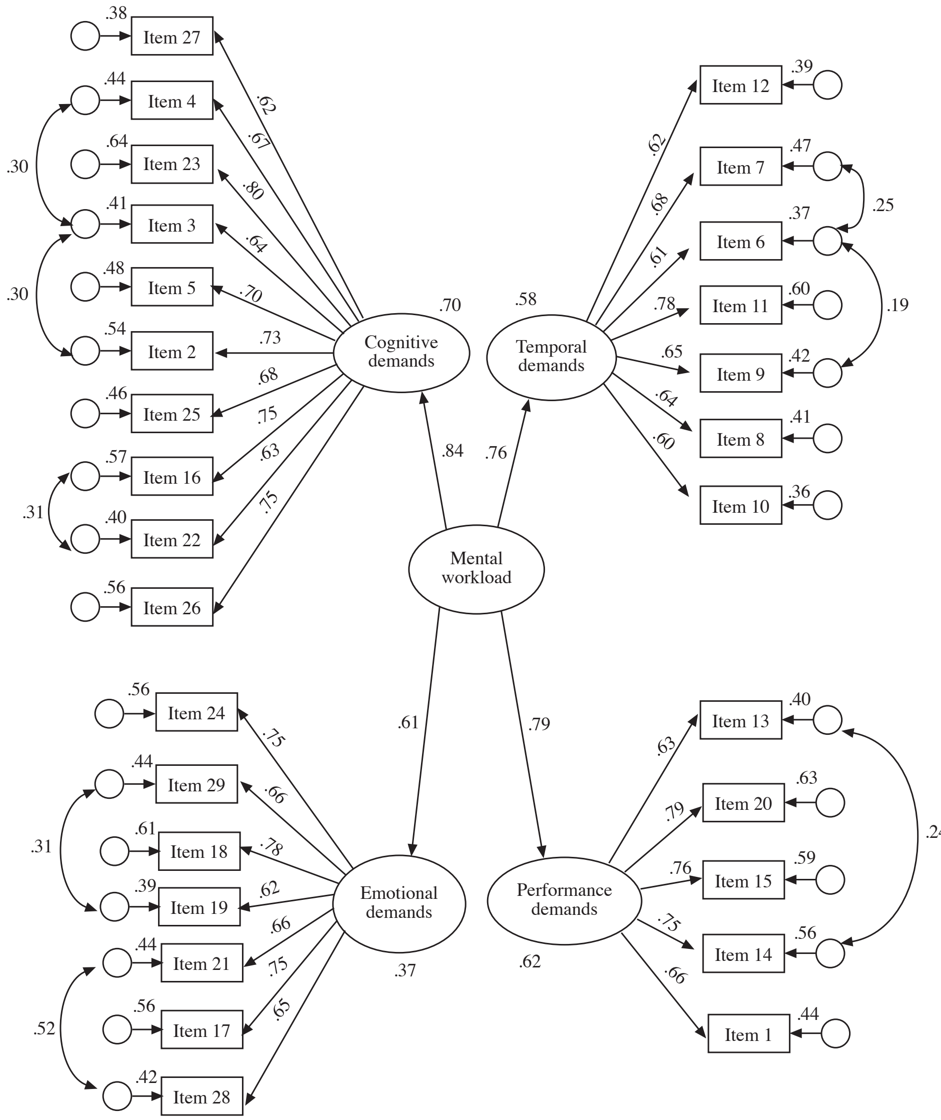

Figure 1. Results of FCA. Standardized solution

Discussion

This study assess the psychometric properties of the CarMen MW Questionnaire in a large sample of workers of different professional sectors. Results showed a four-dimensional model including aspects related to task demands (cognitive, temporal and performance demands) and subject experience (emotional demands). Cognitive demands dimension refers to attentional, complex information processing and decision-making aspects required by the job. Temporal demands dimension includes aspects related to the pace of work and speed demands. Performance demands dimension takes account for performance requirements and job's responsibility degree and, in contrast with NASA-TLX, it doesn't refer to worker's estimation about performance level that he/she reaches in his/her job. Emotional demand dimension of CarMen-Q includes aspects as the degree by which job makes the worker nervous, anxious or stressed. This last dimension also includes items 21 and 28 which ask about worker's physical fatigue (and not about physical demands), a fact supported by existing research in this area which reveals how physical and emotional aspects are linked (Dewe, 1991; Johnson et al., 2005). CarMen-Q structure is consistent with the most frequently MW conceptualization by terms such as mental and emotional strain, an increase in mental effort that comes from anxiety evoked by cognitive aspects of task (Cain, 2007). Significant covariance between some items errors could permit elimination of some item (e.g. item3, item22, item19, item21 or item13), since its content is similar or related to other (see Figure 1). However, we decided to keep them in order to maintain CarMen-Q internal consistency (see Table 3).

Validity analyses based on relationships with NASA-TLX showed high positive correlations as expected and the physical-demand-free nature of CarMen-Q. Significant correlations between NASA-TLX mental demand and CarMen-Q cognitive and performance dimensions would indicate a greater diagnostic power of CarMen-Q, since this would be able to differentiate

Table 3
Descriptive and internal consistency of CarMen-Q

	Mean	Standard Deviation (SD)	Corrected Item-Total correlation	Cronbach's α if Item deleted	Subscales (α; Mean: SD)
Item 27	1.68	.83	.59	.87	Cognitive demands (.88; 18.68; 5.86)
Item 4	1.70	.85	.62	.87	
Item 23	1.97	.84	.71	.86	
Item 3	2.22	.77	.60	.87	
Item 5	2.22	.80	.69	.86	
Item 2	2.17	.84	.63	.87	
Item 25	1.65	.86	.56	.87	
Item 16	1.72	.84	.60	.87	
Item 22	1.43	.85	.53	.87	
Item 26	1.92	.83	.59	.87	
Item 24	1.11	.76	.62	.81	Emotional demands (.84; 7.83: 3.92)
Item 29	.80	.80	.64	.80	
Item 18	1.12	.71	.56	.82	
Item 19	.81	.84	.56	.82	
Item 21	1.23	.78	.60	.81	
Item 17	1.34	.77	.54	.82	
Item 28	1.38	.80	.59	.81	
Item 12	1.55	.83	.46	.76	Temporal demands (.79; 11.36; 3.85)
Item 7	1.44	.84	.59	.74	
Item 6	1.70	.85	.53	.75	
Item 11	1.94	.75	.53	.75	
Item 9	1.39	.80	.48	.76	
Item 8	1.55	.83	.47	.76	
Item 10	1.78	.89	.50	.76	
Item 13	2.37	.73	.60	.75	Performance demands (.80; 11.68; 2.77)
Item 20	2.33	.74	.62	.75	
Item 15	2.14	.88	.64	.74	
Item 14	2.20	.74	.58	.76	
Item 1	2.62	.56	.49	.79	

Total Scale: $\alpha = .905$; *Mean* = 49.55; *SD* = 12.23

between two sources of MW that in NASA-TLX are hidden under a single dimension. Worker's responsibility level and severity of failure's consequences are recognized as MW factors different from those related to the quantity and quality of the information to be handled in the workplace (International Standardization Organization, 1991).

Analysis of CarMen-Q's internal consistency showed appropriate values according to standard recommendations. Values of CR and AVE coefficients confirmed this conclusion. A review of NASA-TLX reliability shows lower internal consistency than obtained here for CarMen-Q. Cronbach's alpha for NASA-TLX is usually around .70 (Changxiu, Xuqun, & Chenming, 2017).

One limitation of the present study is the use of an incidental sample even though the sample was large and representative of a wide range of workers. External validity of CarMen-Q, including criteria related to organizational problems (stress, accidents, sick leave…) that can be affected by MW, is not reported but it will be carried out in future. Another line of future research is to study the presence of differences in MW scores between professional sectors, jobs and sociodemographic characteristics of workers and to carry out the standardisation of CarMen-Q.

In spite of limitations, this study provides evidence of factor structure and internal consistence of CarMen-Q, and may be helpful in search for a clearer definition of the construct. Despite being the most used tool to evaluate subjective MW, no research has been done on NASA-TLX factorial structure nor has its validity been verified in real working environments (Young et al., 2015). Although NASA-TLX has multiple advantages (widely accepted, a quick and easy method of estimating multidimensional MW and easy to apply), also presents some disadvantages as potential response bias and psychometric properties tested only in experimental or laboratory environments but never in real workplaces (Rubio et al., 2004).

MW is one of the most important ergonomic risk factors. The European Pact for Mental Health and Welfare (European Commission, 2008) recognizes the need to conduct MW assessments to promote physical and mental wellbeing. To achieve this goal, CarMen-Q is an easy, short-time consuming and useful tool for occupational health practice and job design.

References

Bian, H. (2011). Structural Equation Modelling with Amos II. Office for the Faculty of Excellence. Retrieved from https://www.scribd.com/doc/283895591/SEMwith-AMOS-II-pdf

Cain, B. (2007). *A review of the mental workload literature*. Defence Research and Development Toronto (Canada). Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a474193.pdf

Changxiu, S., Xuqun, Y., & Chenming, Z. (2017). Analysis of mental workload in online shopping: Are augmented and virtual reality consistent? *Frontiers in Psychology*, *8*, 71. doi: 10.3389/fpsyg.2017.00071

Chiorri, C., Garbarino, S., Bracco, F., & Magnavita, N. (2015). Personality traits moderate the effect of workload sources on perceived workload in flying column police officers. *Frontiers in Psychology, 6*, 1835. doi: 10.3389/fpsyg.2015.01835

Dewe, P. (1991). Primary appraisal, secondary appraisal and coping: Their role in stressful work encounters. *Journal of Occupational Psychology, 64*, 331-51. doi: 10.1111/j.2044-8325.1991.tb00564.x

DiDomenico, A., & Nussbaum, M. A. (2011). Effects of different physical workload parameters on mental workload and performance. *International Journal of Industrial Ergonomics*, *41*, 255-260. doi: 10.1016/j.ergon.2011.01.008

Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing and T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 3-25). London: LEA.

European Commission (2008). *European Pact for health and mental health*. Retrieved from http://ec.europa.eu/health/mental_health/policy_es

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*, 39-50.

Goh, J., Pfeffer, J., & Zenios, S. A. (2015). Workplace stressors & health outcomes: Health policy for the workplace. *Behavioral Science & Policy, 1*, 43-52. doi: 10.1353/bsp.2015.0001

González, J. L. (2003). *Carga mental y fatiga en el trabajo. Análisis en función de variables de personalidad en una muestra de servicios* *especiales de enfermería* [Mental workload and fatigue at work. Analysis based on personality variables in a sample of special services nursing] (Doctoral dissertation). Universidad Autónoma de Madrid: Madrid.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*, 904-908.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139-183.

International Standardization Organization (1991). *Ergonomic principles related to mental workload. General terms and definition, ISO 10075*. Geneva: International Standardization Organization.

Johnson, S., Cooper, C., Cartwright, S., Donald, I., Taylor, P., & Millet, C. (2005). The experience of work-related stress across occupations. *Journal of Managerial Psychology*, *20*, 178-187. doi: 10.1108/02683940510579803

Moreno, B., & Báez, C. (2010). Factores y riesgos psicosociales: formas, consecuencias, medidas y buenas prácticas [Psychosocial factors and risks: forms, consequences, measures and good practices]. Madrid: UAM. Retrieved from http://www.insht.es/InshtWeb/Contenidos/Documentacion/PUBLICACIONES%20PROFESIONALES/factores%20riesgos%20psico.pdf

Nunnally, B. H., & Bernstein, J. C. (1994). *Psychometric Theory*. London, UK: McGraw-Hill.

Rolo-González, G., Díaz-Cabrera, D., & Hernández-Fernaud, E. (2009). Development of a Subjective Mental Workload Scale (SCAM). *Revista de Psicología del Trabajo y de las Organizaciones*, *25*, 29-37.

Rubio, S., Díaz, E., Martín, J., & Puente, J. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and Workload Profile methods. *Applied Psychology: An International Review, 53*, 61-86. doi: 10.1111/j.1464-0597.2004.00161.x

Stansfeld, S., & Candy, B. (2006). Psychosocial work environment and mental health—a meta-analytic review. *Scandinavian Journal of Work, Environment & Health, 32*, 443-462. doi: 10.5271/sjweh.1050

Szalma, J. L. (2008). Individual differences in stress reaction. In P. A. Hancock and J. L. Szalma (Eds.), *Performance under Stress* (pp. 323-357). Hampshire: Ashgate.

Tabachnick, B., & Fidell, L. (2013). *Using Multivariate Statistics*. Boston: Pearson.

Verwey, W. B. (2000). On-line driver workload estimation. Effects of road situation and age on secondary task measures. *Ergonomics, 43*(2), 187-209.

Yeh, Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors*, *30*, 111-120.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, *58*, 1-17. doi: 10.1080/00140139.2014.956151