Clinical and Epidemiologic Research

The Computer-Vision Symptom Scale (CVSS17): Development and Initial Validation

Mariano González-Pérez,[1] Rosario Susi,[2] Beatriz Antona,[1] Ana Barrio,[1] and Enrique González[1]

[1]Faculty of Optics and Optometry, Universidad Complutense de Madrid, Madrid, Spain
[2]Faculty of Statistical Studies, Universidad Complutense de Madrid, Madrid, Spain

PURPOSE. To develop a questionnaire (in Spanish) to measure computer-related visual and ocular symptoms (CRVOS).

METHODS. A pilot questionnaire was created by consulting the literature, clinicians, and video display terminal (VDT) workers. The replies of 636 subjects completing the questionnaire were assessed using the Rasch model and conventional statistics to generate a new scale, designated the Computer-Vision Symptom Scale (CVSS17). Validity and reliability were determined by Rasch fit statistics, principal components analysis (PCA), person separation, differential item functioning (DIF), and item–person targeting. To assess construct validity, the CVSS17 was correlated with a Rasch-based visual discomfort scale (VDS) in 163 VDT workers. Test–retest reliability (two-way single-measure intraclass correlation coefficients [ICC] and their 95% confidence intervals, and coefficients of repeatability [COR]).

RESULTS. The CVSS17 contains 17 items exploring 15 different symptoms. These items showed good reliability and internal consistency (mean square infit and outfit 0.88–1.17, eigenvalue for the first residual PCA component 1.37, person separation 2.85, and no DIF). Pearson's correlation with VDS scores was 0.60 ($P < 0.001$). Intraclass correlation coefficient for test–retest reliability was 0.849 (95% confidence interval [CI], 0.800–0.887), and COR was 8.14.

CONCLUSIONS. The Rasch-based linear-scale CVSS17 emerged as a useful tool to quantify CRVOS in computer workers.

The study involves the development and initial validation process of the first Rasch-based scale designed to assess computer-related ocular and visual symptoms.

Keywords: computer, scale, asthenopia, questionnaire, VDT

Resumen
PROPÓSITO: Desarrollar una escala para medir los síntomas visuales y oculares (CRVOS) asociados al uso de videoterminales (VDT) en el trabajo: La escala CVSS17.
MÉTODOS: Se desarrolló un cuestionario piloto siguiendo el procedimiento recomendado. 636 sujetos lo completaron, y se evaluaron sus respuestas según el modelo de Rasch y estadísticas convencionales para crear el CVSS17. La validez y fiabilidad fueron evaluados mediante el ajuste al modelo de Rasch, el análisis de componentes principales (PCA), el índice de separación para los sujetos, el "funcionamiento diferencial de los ítems" (DIF) y el ajuste entre la dificultad de los ítems y la habilidad de los sujetos. Para evaluar la validez de constructo, el CVSS17 se correlacionó con una escala de molestias visuales [VDS] en 163 usuarios de VDT. La fiabilidad test–retest (coeficiente de correlación intraclase [ICC] con su intervalo de confianza del 95% y coeficiente de repetibilidad [COR]).
RESULTADOS: Los 17 ítems del CVSS17 investigan 15 síntomas diferentes, han demostrado buena fiabilidad y consistencia interna (Infit y Outfit en el intervalo [0.88–1.17], el autovalor del primer contraste del análisis PCA de los resultados era 1,37, la separación para los sujetos era 2.85; y no había DIF). El coeficiente de correlación de Pearson con la VDS fue 0.60 ($P < 0.001$). El ICC fue 0.849 (IC al 95%, 0.800–0.887) y el COR 8,14.
CONCLUSIÓN: El CVSS17 es un instrumento basado en el modelo Rasch, que proporciona una escala lineal apropiada para medir el nivel de CRVOS en trabajadores usuarios de VDT.

Computer-related visual and ocular symptoms (CRVOS) in persons who spend a large proportion of their working day looking at a video display terminal (VDT) are the most frequently occurring health problems among VDT users.[1] Given the high prevalence of these symptoms,[2–4] it is likely that a VDT worker will at some point need an eye exam to assess symptoms associated with VDT use.

Studies designed to estimate the prevalence of CRVOS have used questionnaires self-completed by the study participants. These investigations have mostly been based on questionnaires

created by the study authors themselves, with some exceptions. In two Italian studies (Fenga et al.[5] and Taino et al.[6]), a questionnaire developed by the Italian Society of Occupational Medicine (SIMLII) was employed. Other studies[3,7,8] have also addressed this issue using the questionnaire by Hayes et al.[9] Despite the use of these scoring systems, however, no data are available on their validity or accuracy. Moreover, there is no uniformity in the scales used in these questionnaires to quantify the frequency and/or intensity of CRVOS. Thus, although most questionnaires have been based on ordinal scales of three to five levels, some authors[10,11] have used a visual analog scale (VAS) to estimate the intensity of symptoms. All these scales use conventional summary scoring, which assumes that equal distances between response categories render equal distances in the dimension measured, and that all items represent the same level of difficulty and should consequently be scored equally.[12]

As far as we are aware, no subjective tool designed to evaluate CRVOS has been developed by means of Rasch analysis. This method is recommended for the creation of this type of instrument[13] because (1) it generates a more accurate score, overcoming the limitations of traditional summary scoring through the transformation of ordinal raw scores into interval linear scales[12–15]; and (2) it provides insight into the internal consistency of the scale and is able to match item difficulty to user skill.[15] The Rasch method also provides data like person and item reliability, indicating the overall performance of the instrument.[13]

In this study, we develop a valid, reliable questionnaire in Spanish for the assessment of visual and ocular symptoms, capable of measuring CRVOS in VDT workers. This instrument is comparable to existing convergence insufficiency[16,17] or vision-related quality of life[12,15,18,19] questionnaires.

METHODS

The new questionnaire, Computer-Vision Symptom Scale (CVSS17), was developed following the recommendations of other authors[13,20,21] such that items were generated through qualitative research and then selected and scored by Rasch analysis.

The study was approved by the Research Ethics Committee of the Hospital Clínico San Carlos (Madrid, Spain), and the study protocol adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from all study participants once the nature and possible consequences of the study had been explained to them.

Item Generation and Selection

Items judged appropriate for a CRVOS questionnaire were identified in different ways:

1. Through a search of the different databases (MEDLINE, EMBASE, and PROQOLID) focusing on studies conducted to date on CRVOS[9,22–28];
2. By asking 14 optometrists (with 9 ± 6 years of clinical experience) to detail the words used by their patients to describe these symptoms and list the most common VDT-related complaints;
3. According to the recommendations of others,[13,20,21,29] we conducted semistructured interviews with 59 VDT workers (mean age 38.6 ± 9.2 years, 52.5% female) fulfilling the definition of "VDT worker" established by the Instituto Nacional de Seguridad e Higiene en el Trabajo (INSHT, Spanish Institute of Health and Safety at Work)[30];

4. Also through incorporation of five items of the VFQ25[31] and one item of the VF14[32] questionnaires.

In this first stage, we obtained a pool of 277 items. Two optometrists then used an item assessment guide based on the recommendations of Streiner and Norman[21] (see Supplementary Table S1 for details) to reduce the item bank to 138. These 138 items were evaluated by a group of 16 volunteer users who were instructed to choose the items that best described each symptom. In addition, for each proposed item they chose the response category group, among the groups used in similar questionnaires cited in the literature, that best described the severity of the symptoms they experienced at work.

This process served to generate 77 items for a pilot questionnaire fulfilling the following inclusion criteria: There had to be at least one item for each symptom described in the prior item-generation stages; if a user's preferred item for a symptom differed from the item best rated by the experts, both were included. Also, the response category group for each item was chosen by the users in such a manner that initially one item had a seven-category response scale, 34 had a six-category scale, 26 had a five-category scale, and 16 had a four-category scale.

Pilot Questionnaire

The pilot questionnaire (CVSS77) consisted of the 77 items selected as described above plus 11 items designed to obtain information on age (18–65 years), sex, and whether the respondent fulfilled the criteria for a "VDT worker" as defined by the Spanish INSHT.[30] Subjects were required to provide replies for at least 66% of all items.

The pilot CVSS77 was distributed among the members of a trade union (Unión General de Trabajadores) and a health and safety at work organization (Grupo OTP-Prevención de Riesgos Laborales) from May 7 to October 19, 2012 via their Web sites. Each time the Web site was accessed, one of six versions of the questionnaire with the items in different order appeared to avoid order effects.

The questionnaire was completed online by 636 subjects. Forty-eight questionnaires were eliminated because they were incorrectly completed, leaving 588 completed questionnaires for validation.

The Rasch model is an item response theory (IRT) model. The model transforms raw scores to preserve the distance between the locations of two persons regardless of the particular items administered. The main IRT concept is that a mathematical model is used to predict the probability of a person successfully replying to an item according to the person's ability and item difficulty.[33]

Since the selected items were polychotomous, for Rasch analysis we had to choose between the partial credit model (PCM, which considers a different rating scale for each item) and the Andrich rating scale model (RSM, which assumes equal category thresholds across items). The PCM is less restrictive than RSM because it allows for different response categories in different items, yet it may complicate the communication to the audience and requires a larger dataset.[34] The PCM was finally selected for two reasons: (1) RSM would mean making a priori assumptions about the similarity of scale points across items, and we had no evidence of this in our item set; and (2) several items (e.g., A30–A22 and B7–B8) initially showed different response patterns despite sharing the same rating scale structure, so PCM was likely to offer more scoring precision than RSM.

The PCM implemented in BIGSTEPS software (version 2.82) was used to identify unusual response patterns. Infit and outfit mean square values, which compare predicted and observed responses, were obtained for each subject and, according to
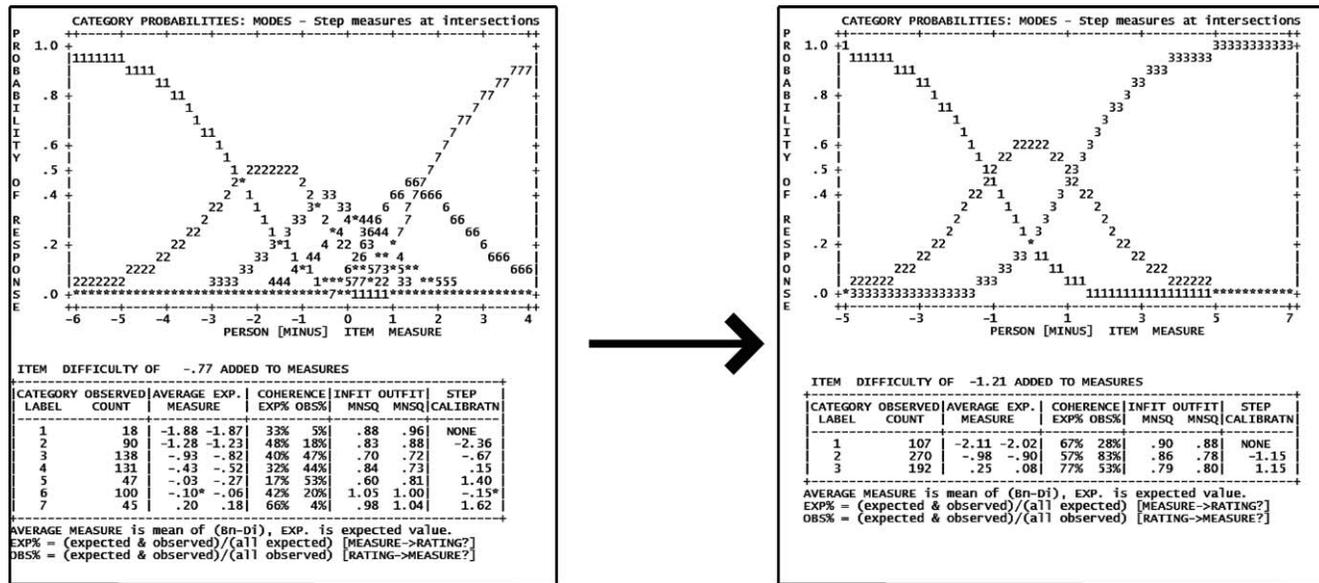
//xinet/production/i/iovs/live_jobs/iovs-55-06/iovs-55-06-50/layouts/iovs-55-06-50.3d ∎ 25 June 2014 ∎ 8:59 pm ∎ Allen Press, Inc. ∎ Customer #IOVS-13-13818

Page 2

FIGURE 1. A4 item characteristic curves. The curve on the *left* shows disordered thresholds and does not fulfill some of the criteria recommended by Linacre.[36] Consequently, categories initially scored as 5, 6, and 7 were rescored as 3; categories initially scored 3 and 4 were rescored as 2; and category initially scored as 2 was rescored as 1. The resultant curve (*right*) shows no disordered thresholds and fulfills the established criteria.

established criteria,[35] four questionnaires were revised because their outfit was >2.5; two of these were discarded because responses lacked coherence. This left 586 valid completed questionnaires. A further 10 questionnaires were excluded by BIGSTEPS because scores were under the minimum estimated measure, leaving 576 valid responses.

Optimizing Rating Scales

A productive early step when analyzing questionnaire and survey data is to assess the functioning of rating scale categories.[36] Hence, we assessed the performance of the initial response scales according to the eight criteria recommended by

Linacre[36]: at least 10 observations for each category; a regular observation distribution; average measures should increase monotonically with category; the outfit of each category should be less than 2.0; the responses of subjects with a higher level of symptoms should correspond to the higher rating scale categories; ratings should imply measures, and measures should imply ratings; step difficulty increments should be at least 1.4 logits; these increments should be less than 5.0 logits. After applying these criteria, the initial number of response categories (seven categories in one item, six in 34, five in 26, and four in 16) was markedly reduced from 405 to 227. The final composition of the CV77 was thus 15 items with two response categories, 51 with three, and 11 with four. Owing to this
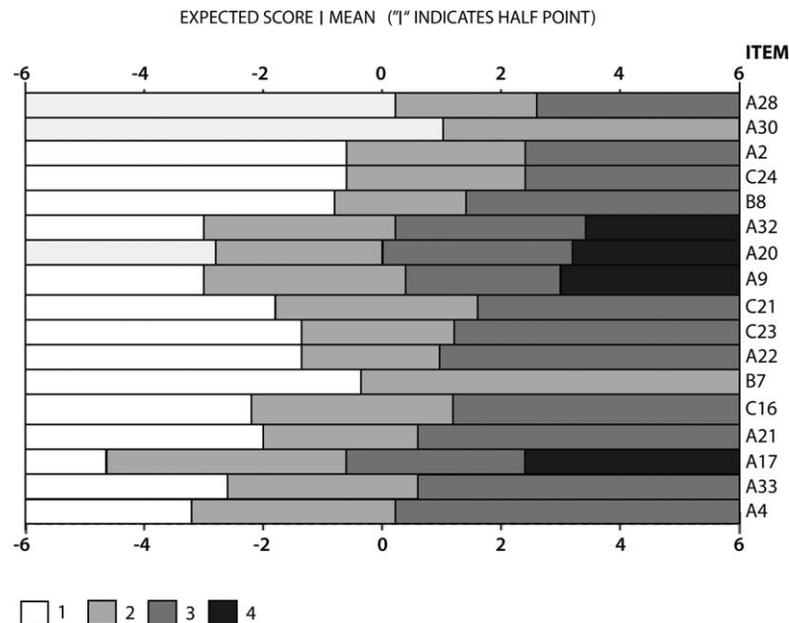

FIGURE 2. Graph showing the most likely category a person with a given severity of symptoms (expressed in logits) would choose as a response to the item shown on the *right*. Rating information is shown in terms of expected scores (1 indicates a score of half a point). The lower rating categories appear on the *left*. Items are ordered from most (*top*) to least (*bottom*) difficult.

TABLE 1. Psychometric Properties of the CVSS17

Item	Skew	Kurtosis	Ceiling Effect	Missing Data	PTBIS. CORR	Measure	Error	MNSQ	ZSTD	MNSQ	ZSTD
A2	.833	−.324	6.94%	0.17%	0.55	0.97	0.09	1.02	0.30	1.04	0.40
A4	−.243	−.945	33.85%	1.04%	0.65	−1.49	0.08	0.91	−1.80	0.88	−2.10
A9	.353	−.092	4.17%	0.17%	0.66	0.12	0.08	0.90	−1.80	0.89	−1.90
A17	.214	−.215	7.12%	0.35%	0.66	−0.92	0.08	0.89	−1.90	0.89	−1.90
A20	.079	−.750	3.47%	0.52%	0.57	0.17	0.07	1.16	2.60	1.17	2.70
A21	.080	−1.291	26.91%	0.52%	0.67	−0.71	0.08	0.92	−1.60	0.88	−1.80
A22	.424	−1.192	20.31%	0.87%	0.65	−0.11	0.08	0.93	−1.20	1.03	0.30
A28	1.336	.750	5.73%	0.17%	0.48	1.39	0.09	1.08	1.20	1.19	1.10
A30	2.367	3.617	22.13%	0.35%	0.43	0.99	0.12	1.03	0.60	0.92	−0.40
A32	.141	−.464	2.60%	0.69%	0.66	0.27	0.08	0.90	−1.80	0.89	−2.00
A33	−.029	−.950	25.35%	0.52%	0.62	−0.95	0.08	0.97	−0.50	1.00	0.00
B7	.325	−1.901	41.49%	1.22%	0.48	−0.34	0.10	1.08	1.60	1.13	1.10
B8	.740	−.817	15.28%	1.04%	0.59	0.38	0.08	1.06	0.90	1.10	0.80
C16	.156	−.928	17.53%	8.16%	0.59	−0.44	0.08	1.04	0.80	1.02	0.40
C21	.271	−.811	13.02%	8.85%	0.63	−0.07	0.09	0.94	−1.00	0.96	−0.60
C23	.403	−1.124	17.88%	6.42%	0.59	−0.09	0.08	1.11	1.80	1.03	0.30
C24	.721	−.461	5.90%	14.41%	0.55	0.83	0.09	1.08	1.20	1.17	1.60

modification, the infit for persons improved from 1.09 to 1.00 (best mean square fit statistic = 1), the outfit for persons from 1.17 to 1.02, the infit for items from 1.02 to 1.01, and the outfit for items from 1.21 to 1.03. As an example, Figure 1 shows the transformation from a seven-category item to a three-category item.

Item Reduction

A questionnaire should contain the smallest number of items needed to maintain the best reliability and validity. To pursue this goal, we followed the recommendations of several authors[13,33] in the following order:

- Removal of items showing point-biserial correlations above 0.4.
- Removal of items showing infit and outfit mean square between 0.8 and 1.2.
- Items with means farthest from the subject mean were considered for removal.
- Items showing a high proportion of missing data (>50%) were considered for removal.
- Items showing a high proportion of end category responses (>50%) were considered for removal.
- Items showing a considerably different standard deviation of scores from other items were considered for removal.
- Items with a pattern of response far from normal, with coefficients of skewness and kurtosis outside the range +2 to −2, were considered for removal.

These criteria were applied in an iterative manner (one at a time) to give a 36-item version of the questionnaire (CVSS36).

Differential Item Functioning Analysis (DIF)

Items of the CVSS36 were checked to ensure that there was no difference in the way subgroups (male–female; presbyopes–nonpresbyopes) responded to each item, that is, no DIF using Jmetrik 2.1.0. In this test, the Mantel-Haenszel statistic and its associated P value, effect size (ES), and confidence intervals at 95% are calculated. Subjects who replied that they were over 39 years of age were considered presbyopes.

To obtain a DIF-free questionnaire, three criteria were iteratively applied. Items showing DIF between age and/or sex groups were the first candidates for deletion; among these, we deleted those showing the greater difficulty (to adjust the difficulty of the questionnaire to the subjects, for which Rasch-derived person and item statistics were obtained) and had a Tau-b correlation above 0.5. This process left 17 items free from DIF. These items were thus used to create the CVSS17 questionnaire.

Unidimensionality, Validity, and Reliability of the CVSS17

The unidimensionality of the CVSS 17 was determined by principal components analysis (PCA) of standardized residuals using BIGSTEPS. Multidimensionality is assumed when the first contrast has the power of at least two items (eigenvalue > 2.0).[15]

CVSS17 validity was assessed through Rasch analysis, factor analysis, and Cronbach's α.

An initial clinical validation was done by comparing, in a group of 163 subjects fulfilling the "VDT worker" definition, their CVSS17 and their visual discomfort scale (VDS)[36] (Spanish version, provided as Supplementary Material) scores. In addition, the data collected for these 163 workers, who completed the CVSS17 on two separate occasions, were used to assess the tool's test–retest reliability. For this purpose, after establishing the normality of scores, the two-way single-measure intraclass correlation coefficient (ICC) with its confidence interval (CI) at 95% was calculated. In addition, Bland-Altman limits of agreement were determined to calculate the coefficient of repeatability (COR) by subtracting the mean difference from the upper 95% limit.[37]

RESULTS

For the final 576 study participants (age: 42.8 ± 10.1 years; 54.2% female; 39.4% nonpresbyopes), the mean score recorded was 30.9, median was 30.8, minimum was 18.0, maximum was 50.0, and standard deviation was 7.5. The 95% CI for the population mean was 30.29 to 31.51.

CVSS17 Rating Scale

The CVSS17 contains 17 items with different rating scales. Two items (A30 and B7) have two response categories; 11 items

?8

TABLE 2. Summary Statistics for Persons and Items

INPUT: 588 PERSONS, 77 ITEMS   ANALYZED: 576 PERSONS, 17 ITEMS, 53 CATS

SUMMARY OF   576 MEASURED (NON-EXTREME) PERSONS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	30.2	16.5	-0.89	0.50	0.99	-0.20	1.01	-0.10
S.D.	7.4	0.7	1.69	0.12	0.44	1.20	0.58	1.10
MAX.	50.0	17.0	3.46	1.10	2.99	3.80	5.12	3.70
MIN	17.0	13.0	-5.12	0.42	0.23	-3.20	0.13	-3.20

REAL RMSE.	0.56	ADJ.SD	1.60	SEPARATION	2.85	PERSON RELIABILITY		0.89
MODEL RMSE.	0.52	ADJ.SD	1.61	SEPARATION	3.10	PERSON RELIABILITY		0.91

S.E. OF PERSON MEAN   0.07

WITH 10 EXTREME PERSONS = 586   PERSONS MEAN -0.97   S.D. 1.80

REAL RMSE.	0.59	ADJ.SD	1.70	SEPARATION	2.89	PERSON RELIABILITY		0.89
MODEL RMSE.	0.55	ADJ.SD	1.71	SEPARATION	3.12	PERSON RELIABILITY		0.91

MINIMUM EXTREME SCORE: 10 PERSONS
DELETED: 2 PERSONS
VALID RESPONSES: 97.3%

SUMMARY OF 17 MEASURED ITEMS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	1021.8	506.6	0.00	0.08	1.00	0.00	1.01	-0.10
S.D.	199.6	23.3	0.75	0.01	0.08	1.50	0.11	1.40
MAX.	1414.0	575.0	1.39	0.12	1.16	2.60	1.19	2.70
MIN	701.0	493.0	-1.49	0.07	0.89	-1.90	0.88	-2.10

REAL RMSE.	0.09	ADJ.SD	0.75	SEPARATION	8.61	ITEM RELIABILITY		0.99
MODEL RMSE.	0.09	ADJ.SD	0.75	SEPARATION	8.78	ITEM RELIABILITY		0.99

S.E. OF ITEM MEAN   0.19
DELETED: 60 ITEMS

Score is the raw score.

Count is the number of responses made.

Measure is the estimated measure (for persons) or calibration (for items).

Error is the standard error of the estimate.

For infit: MNSQ is the mean square infit statistic with expectation 1. Values substantially below 1 indicate dependency in the data; values substantially above 1 indicate noise; ZSTD is the infit mean square fit statistic standardized to approximate a theoretical mean 0 and variance 1 distribution.

For outfit: MNSQ is the mean square outfit statistic, with expectation 1. Values substantially less than 1 indicate dependency in the data; values substantially greater than 1 indicate the presence of unexpected outliers; ZSTD is the outfit mean square fit statistic standardized to approximate a theoretical mean 0 and variance 1 distribution.

RMSE is the root mean square standard error computed over the persons or over the items. Model RMSE is computed on the basis that the data fit the model, and that all misfit in the data is merely a reflection of the stochastic nature of the model. This is a "best case" reliability, which reports an upper limit to the reliability of measures based on this set of items for this sample. Real RMSE is computed on the basis that misfit in the data is due to departures in the data from model specifications. This is a "worst case" reliability, which reports a lower limit to the reliability of measures based on this set of items for this sample.

ADJ.S.D. is the standard deviation of the estimates after subtracting from their observed variance the error variance attributable to their standard errors of measurement.

ADJ.S.D. = SD of measure − RMSE.

The ADJ.S.D. is an estimate of the "true" standard deviation from which the bias caused by measurement error has been removed.

Person (or item) SEP is the ratio of the ADJ.S.D. to RMSE. It provides a ratio measure of separation in RMSE units, which is easier to interpret than the reliability correlation.

Person (or item) SEP REL is a separation reliability, equivalent to KR-20, Cronbach's α, and the generalizability coefficient. The relationship between separation SEP and reliability REL is REL = SEP/(1 + SEP) or SEP = (REL/[1 − REL]).

S. E. of mean is the standard error of the mean of the person (or item) sample.

(A2, A4, A21, A22, A28, A33, B8, C16, C21, C23, and C24) have three response categories; and 4 items have four response categories (A9, A17, A20, and A32). Figure 2 shows the most likely category a person with a given severity of symptoms (expressed in logits) would choose as a response to the item shown on the right.[38]

Psychometric Properties of the CVSS17

The criteria established were met by every item except A30, whose skewness and kurtosis were out of range (+2, −2). Item A30 was not eliminated since we considered that it contributed clinical information not provided by any other of the
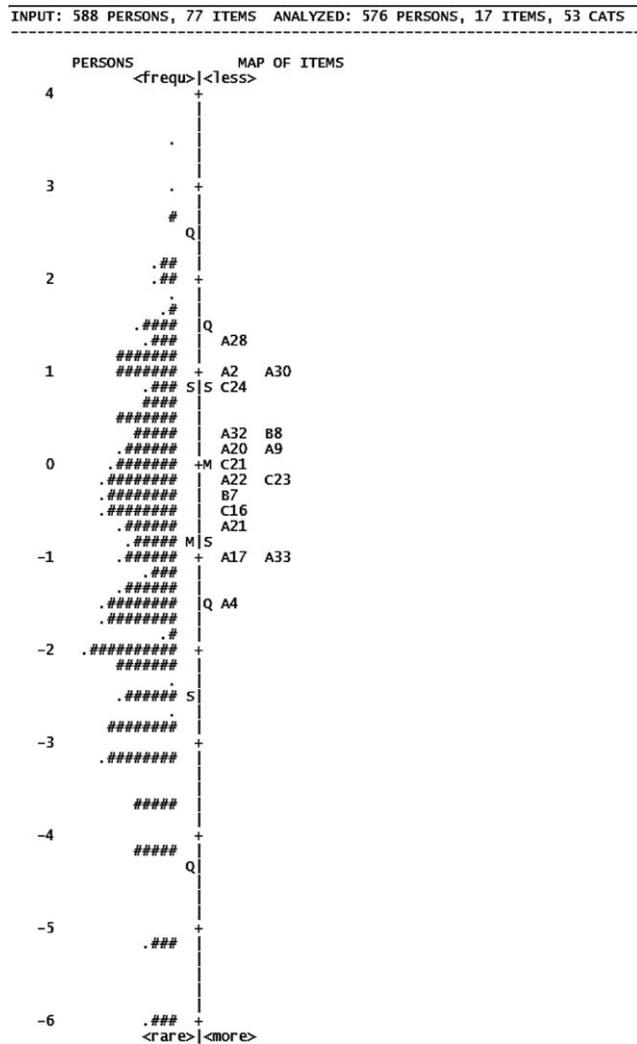
FIGURE 3. Person/item map for the CVSS17 questionnaire. *Left-hand* column indicates personal ability expressed in logits. Measures alongside the items (persons with the more intense symptoms appear on the *top*) and the *right-hand* column indicates the item difficulty level expressed in logits.

questionnaire items, as confirmed by its correlation (Tau-b) with the remaining items of less than 0.4 in all cases. Data quality is provided in Table 1 as skew and kurtosis, missing data percentages, percentage of responses at the most able end category of the response scale (ceiling effect), point-biserial correlation, item difficulty (measure), and infit/outfit mean square statistics for each item of the CVSS17.

Table 2 summarizes the data generated for the CVSS17 including infit, outfit, and separation and person reliability indices, among others. As may be seen in the table, the difference between the average difficulty of the items and subjects was −0.89 logits. This difference can be seen in greater detail in Figure 3.

Factor analysis and varimax rotation revealed two main components: factor 1 including items A4, A9, A17, A20, A21, A32, B7, B8, C16, and C23 and factor 2 comprising A2, A22, A28, A30, A33, C21, and C24. Figure 4 shows factor loadings for each item. There was no DIF for the age and/or sex groups in any item.

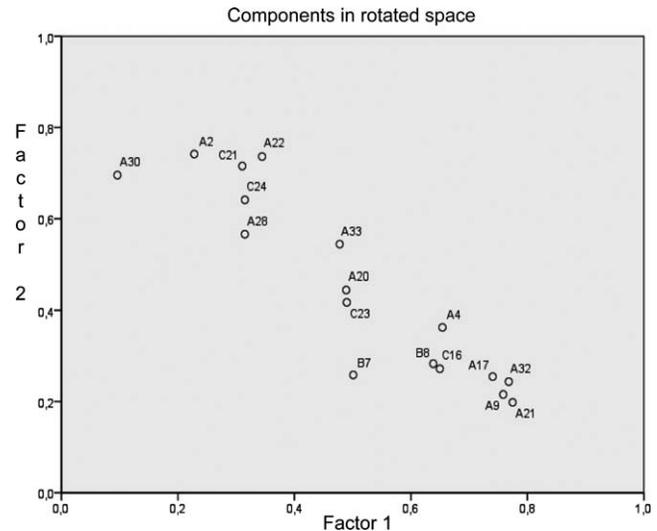Rasch-based PCA for score residuals using BIGSTEPS returned an eigenvalue for the first component below 2.0

FIGURE 4. Scatter plot of loadings for the main components in rotated space.

(1.37). This confirmed the hypothesis that residuals are random noise. Cronbach's α was 0.92.

CVSS17 Performance

The Kolmogorov-Smirnov test (K-S) for the CVSS17 and VDS scores obtained indicated a normal distribution for both measures ($P > 0.05$). Accordingly, we calculated Pearson's correlation coefficients between the total scores for CVSS17 and VDS at 0.60: 0.51 ($P < 0.001$) for VDS–factor 1 correlation and 0.57 ($P < 0.001$) for VDS–factor 2 correlation.

In the subjects who completed the CVSS17 twice (test–retest time interval: 15.83 ± 3.20 days), two-way, single-measure ICC for test–retest reliability was 0.849 (95% CI, 0.800–0.887), and the COR was 8.14. Figure 5 shows the Bland-Altman plot for CVSS17.
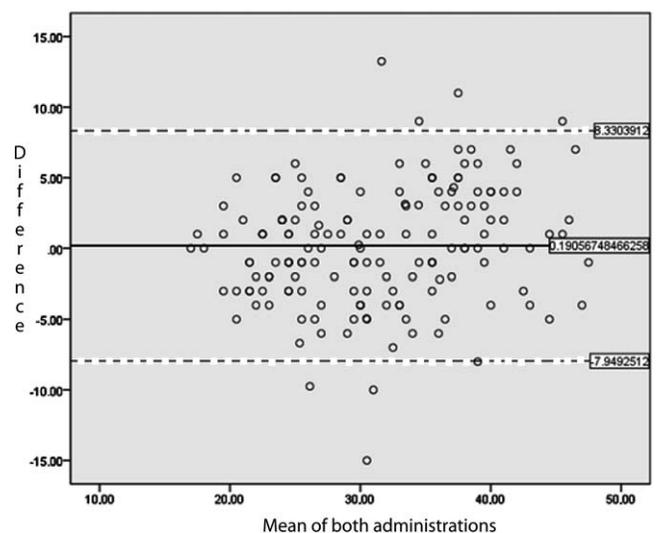

FIGURE 5. Plot of mean differences versus repeatability of the CVSS17 scores. The *solid line* indicates the mean difference (MD) between scores obtained when completing the questionnaire on two occasions. The *dotted lines* indicate the lower and the upper 95% limits of agreement (MD ± 1.96 × SD).

?13

?14

DISCUSSION

This paper describes a new tool to quantify vision-related symptoms associated with VDT use at work, developed using conventional techniques and Rasch analysis to provide reliable and valid measures.

The final number of items included (17) is similar to those of other available vision-related validated questionnaires.[12,15,16,39] This number of items means that the subject can complete the questionnaire quickly, especially if in electronic format.

The 17 items of the scale were designed to obtain information about 15 different symptoms. These symptoms have been included with different frequencies in other questionnaires used in research on CRVOS.[6,9-11] The behavior of the symptoms defined in our questionnaire resembles that of the two main contributing factors in the factorial analysis described by Sheedy et al.[40] for experimentally induced asthenopia. However, the CVSS17 includes a broader range of symptoms like photophobia (A33 and C23) and "blinking a lot" (A20), which were noticeably influenced by these two factors. The detection of two main factors, one related to the external symptom factor of Sheedy et al.[40] and the other related to the internal symptom factor, along with the presence of photophobia, suggests that the symptom model assessed by CVSS17 is similar to existing described models.[9-11,40]

The item identification and reduction methods used in the CVSS17 development were systematic and rigorous in order to ensure content validity.[13,14,41] The PCM was used to reorder the response categories. This enables the selection of items with a good discrimination capacity and provides a statistically justified scale, without significant missing data, that shows ordered thresholds on Rasch analysis.

Because the selected items had different question formats (i.e., symptom severity, symptom frequency, subject opinion), we decided to include several rating scales that were chosen by a set of study subjects according to their suitability. The aim was, as far as possible, to use the most appropriate rating scale for each item. However, based on recently published data,[42,43] we consider the use of multiple rating scales as the major limitation of the CVSS17 because they can increase respondent burden, and also because they provided some evidence that differences in rating scale formats have some effect on an item's calibrations beyond item content.[43] Although the measurement properties of the CVSS17 may not be compromised per se, this should be taken into account when one is interpreting its item difficulty estimates, investigating improvements to the instrument, and comparing CVSS17 scores with similar scales.

The Rasch statistics used revealed that all items fit the model and, together with the residual PCA, confirmed its unidimensionality. Moreover, the point-biserial correlation calculated for each item of the CVSS17 was in the range 0.43 to 0.67, indicating significant yet nonredundant correlation.

Although Cronbach's α coefficient is not considered a useful measure of the reliability of a scale,[13] we decided to include it in our analysis to facilitate comparisons with other scales. For clinical applications, a coefficient between 0.9 and 0.95 is recommended.[44] Thus, we consider that the internal consistency of the CVSS17 (Cronbach's α = 0.92) makes it useful for comparisons between groups and for clinical applications.

The person separation value obtained (2.85) indicates that the tool is sufficiently sensitive to distinguish between high and low performers, and the person reliability index (0.89) indicates a capacity of CVSS17 to distinguish three or four levels of symptoms.[45] Also, the item separation value calculated (8.61) indicates that the person sample was large

enough to confirm the item difficulty hierarchy (i.e., construct validity) of the tool.[45]

The final questionnaire showed no DIF for the defined groups (male–female, presbyopes–nonpresbyopes). This means there was no difference in the way in which these subgroups responded to the test, indicating the validity of CVSS for all these subgroups.[33]

The mean difference between person capacity and item difficulty was −0.89 logits, a little over the 0.5 logits difference recommended by Pesudovs et al.,[13] indicating that items targeted the more symptomatic end of the CVSS17. This is common for a symptom scale due to the presence in the sample of many subjects with few or no symptoms[15] and/or to a tendency for subjects to underreport their discomfort.

The summary statistics of the Rasch model confirmed that all the selected items contribute significantly to the overall score and that they all measure a related concept. Based on these observations, we propose that this concept is the set of visual and ocular symptoms associated with work-time VDT use.

Given the lack of a gold standard with which to compare our CVSS17 data, we used another validated instrument that measures a closely related concept, the VDS,[46] which has been used to measure reading-related visual discomfort.[47] Significant moderate to high correlation was detected between this scale and CVSS17, and it also correlated significantly with the two main factors of our scale. For factor–VDS correlations, it was bigger for factor 2. These correlates can be considered the first evidence of the validity of CVSS17. **?9**

According to ICCs, test–retest reliability for the CVSS17 was good. The COR was somewhat higher than expected, probably due to the influence of eight subjects whose scores varied by 10 points or more when the questionnaire was completed twice. This was revealed by the fact that the ICC and COR significantly improved when this analysis was repeated with these subjects excluded.

The printed Spanish version of the CVSS17 and its Rasch-based scoring chart are provided as Supplementary Material. We also provide an English version for its potential international use. However, more clinical research is needed to obtain more evidence of the validity of the scale (discriminant validity, divergent validity, and further evidence of construct validity) and to determine normal values of CVSS for population subgroups varying in socioeconomic status, race, and so on. Future studies will also need to determine the extent to which CVSS17 can detect clinically important changes over time (minimum clinically important difference, MID).

In conclusion, the CVSS17 questionnaire was developed using conventional techniques and Rasch analysis, ensuring construct validity and providing measures as a linear interval scale rather than ordinal measures. The CVSS17 is therefore able to assess CRVOS without the main limitations of previously developed instruments.[12,13,15]

References

1. Blehm C, Vishnu S, Khattak A, Mitra S, Yee R. Computer vision syndrome: a review. *Surv Ophthalmol*. 2005;50:253–262.

2. Ostrovsky A, Ribak J, Pereg A, Gaton D. Effects of job-related stress and burnout on asthenopia among high-tech workers. *Ergonomics*. 2012;55:854–862.

3. Portello J, Rosenfield M, Bababekova Y, Estrada J, Leon A. Computer-related visual symptoms in office workers. *Ophthalmic Physiol Opt*. 2012;32:375–382.

4. Sa E, Ferreira Junior M, Rocha L. Risk factors for computer visual syndrome (CVS) among operators of two call centers in São Paulo, Brazil. *Work*. 2012;41(suppl 1):3568–3574.

5. Fenga C, Di Pietro R, Fenga P, et al. Astenopia e lavoro al VDT: nostra esperienza. *G Ital Med Lav Erg*. 2007;29:500–501.

6. Taino G, Ferrari M, Mestad I, Fabris F, Imbriani M. Astenopia e lavoro al videoterminale: studio di una popolazione di 191 lavoratori esposti al rischio mediante somministrazione di questionario anamnestico mirato e valutazione oftalmologica. *G Ital Med Lav Erg*. 2006;28:487–497.

7. Chu C, Rosenfield M, Portello J, Benzoni J, Collier J. A comparison of symptoms after viewing text on a computer screen and hardcopy. *Ophthalmic Physiol Opt*. 2011;31:29–32.

8. Rosenfield M, Hue J, Huang R, Bababekova Y. The effects of induced oblique astigmatism on symptoms and reading performance while viewing a computer screen. *Ophthalmic Physiol Opt*. 2012;32:142–148.

9. Hayes J, Sheedy J, Stelmack J, Heaney C. Computer use, symptoms, and quality of life. *Optom Vis Sci*. 2007;84:738–744.

10. Aakre B, Doughty M. Are there differences between "visual symptoms" and specific ocular symptoms associated with video display terminal (VDT) use? *Cont Lens Anterior Eye*. 2007;30:174–182.

11. Helland M, Horgen G, Kvikstad T, Garthus T, Bruenech J, Aarås A. Musculoskeletal, visual and psychosocial stress in VDT operators after moving to an ergonomically designed office landscape. *Appl Ergon*. 2008;39:284–295.

12. Pesudovs K, Garamendi E, Elliott D. The Quality of Life Impact of Refractive Correction (QIRC) Questionnaire: development and validation. *Optom Vis Sci*. 2004;81:769–777.

13. Pesudovs K, Burr J, Harley C, Elliott D. The development, assessment, and selection of questionnaires. *Optom Vis Sci*. 2007;84:663–674.

14. Khadka J, McAlinden C, Pesudovs K. Quality assessment of ophthalmic questionnaires: review and recommendations. *Optom Vis Sci*. 2013;90:720–744.

15. McAlinden C, Pesudovs K, Moore J. The development of an instrument to measure quality of vision: the Quality of Vision (QoV) questionnaire. *Invest Ophthalmol Vis Sci*. 2010;51: 5537–5545.

16. Borsting E, Rouse M, Mitchell G, et al. Validity and reliability of the revised convergence insufficiency symptom survey in children aged 9 to 18 years. *Optom Vis Sci*. 2003;80:832–838.

17. Rouse M, Borsting E, Mitchell G, et al. Validity and reliability of the revised convergence insufficiency symptom survey in adults. *Ophthalmic Physiol Opt*. 2004;24:384–390.

18. Massof R, Fletcher D. Evaluation of the NEI visual functioning questionnaire as an interval measure of visual ability in low vision. *Vision Res*. 2001;41:397–413.

19. Massof R, Rubin G. Visual function assessment questionnaires. *Surv Ophthalmol*. 2001;45:531–548.

20. Rattray J, Jones M. Essential elements of questionnaire design and development. *J Clin Nurs*. 2007;16:234–243.

21. Streiner D, Norman G. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 3rd ed. New York: Oxford University Press; 2003;. **?10**

22. Cole B, Maddocks J, Sharpe K. Effect of VDUs on the eyes: report of a 6-year epidemiological study. *Optom Vis Sci*. 1996; 73:512–528.

23. Coren S, Hakstian A. Color vision screening without the use of technical equipment: scale development and cross-validation. *Percept Psychophys*. 1988;43:115–120.

24. Coren S, Hakstian A. Validation of a self-report inventory for the measurement of visual acuity. *Int J Epidemiol*. 1989;18: 451–456.

25. Coren S, Hakstian A. Screening for stereopsis without the use of technical equipment: scale development and cross-validation. *Int J Epidemiol*. 1996;25:146–151.

26. Feldman J, Cooper J, Reinstein F, Swiatoca J. Asthenopia induced by computer-generated fusional vergence targets. *Optom Vis Sci*. 1992;69:710–716.

27. McKeon C, Wick B, Aday L, Begley C. A case-comparison of intermittent exotropia and quality of life measurements. *Optom Vis Sci*. 1997;74:105–110.

28. Vertinsky T, Forster B. Prevalence of eye strain among radiologists: influence of viewing variables on symptoms. *Am J Roentgenol*. 2005;184:681–686.

29. Snyder C, Watson M, Jackson J, Cella D, Halyard M. Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Patient-reported outcome instrument selection: designing a measurement strategy. *Value Health*. 2007;10(suppl):76–85.

30. Instituto Nacional de Seguridad e Higiene en el Trabajo. Guía Técnica para la evaluación y prevención de los riesgos relativos a la utilización de equipos con Pantallas de visualización. Madrid: Ministerio de Trabajo e Inmigración; 2006;. **?11**

31. Broman A, Munoz B, West S, et al. Psychometric properties of the 25-item NEI-VFQ in a Hispanic population: Proyecto VER. *Invest Ophthalmol Vis Sci*. 2001;42:606–613.

32. Steinberg E, Tielsch J, Schein O, et al. The VF-14: an index of functional impairment in patients with cataract. *Arch Ophthalmol*. 1994;112:630–638.

33. Wu M, Adams R. *Applying the Rasch Model to Psycho-Social Measurement: A Practical Approach*. Melbourne: Educational Measurement Solutions; 2007:87.

34. Linacre J. Comparing "Partial Credit Models" (PCM) and "Rating Scale Models" (RSM). *Rasch Measurement Transactions*. 2000;14:768.

35. Lamoureux E, Pallant J, Pesudovs K, Rees G, Hassell J, Keeffe J. The impact of vision impairment questionnaire: an assessment of its domain structure using confirmatory factor analysis and rasch analysis. *Invest Ophthalmol Vis Sci*. 2007;48:1001–1006.

36. Linacre J. Optimizing rating scale category effectiveness. *J Appl Meas*. 2002;3:85–106.

37. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1:307–310.

38. Linacre J, Wright BA. *User's Guide to BIGSTEPS: Rasch Model Computer Programs*. Chicago: Mesa Press; 1998;. **?12**

39. Garamendi E, Pesudovs K, Stevens M, Elliott D. The Refractive Status and Vision Profile: evaluation of psychometric properties and comparison of Rasch and summated Likert-scaling. *Vision Res*. 2006;46:1375–1383.

40. Sheedy J, Hayes J, Engle J. Is all asthenopia the same? *Optom Vis Sci*. 2003;80:732–739.

41. Brod M, Tesler L, Christensen T. Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res*. 2009;18:1263–1278.

42. Khadka J, Gothwal V, McAlinden C, Lamoureux E, Pesudovs K. The importance of rating scales in measuring patient-reported outcomes. *Health Qual Life Outcomes*. 2012;10:80.

43. Khadka J, McAlinden C, Gothwal V, Lamoureux E, Pesudovs K. The importance of rating scale design in the measurement of patient-reported outcomes using questionnaires or item banks. *Invest Ophthalmol Vis Sci*. 2012;53:4042–4054.

Investigative Ophthalmology & Visual Science

44. Bland J, Altman D. Statistics notes: Cronbach's alpha. *BMJ*. 1997;314:572.

45. Linacre J. *Winsteps® Rasch Measurement Computer Program User's Guide*. Beaverton, Oregon: Winsteps.com; 2014.

46. Conlon E, Lovegrove W, Chekaluk E, Pattison P. Measuring visual discomfort. *Vis Cogn*. 1999;6:637–666.

47. Borsting E, Chase C, Ridder WR. Measuring visual discomfort in college students. *Optom Vis Sci*. 2007;84:745–751.

Investigative Ophthalmology & Visual Science