PAIN® 152 (2011) 2399–2404

PAIN®

www.elsevier.com/locate/pain

# Validity of four pain intensity rating scales

Maria Alexandra Ferreira-Valente [a,b,c,*], José Luís Pais-Ribeiro [a,c], Mark P. Jensen [d]

[a] Faculdade de Psicologia e Ciências da Educação da Universidade do Porto, Porto, Portugal
[b] Portuguese Foundation for Science and Technology, Lisbon, Portugal
[c] Unidade de Investigação em Psicologia e Saúde (Psychology and Health Unit), Lisbon, Portugal
[d] Department of Rehabilitation Medicine, University of Washington School of Medicine, Seattle, WA, USA

*Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.*

ARTICLE INFO

*Article history:*
Received 22 February 2011
Received in revised form 25 June 2011
Accepted 11 July 2011

*Keywords:*
Pain assessment
Validity
Numerical Rating Scale
Visual Analogue Scale
Faces Pain Scale
Verbal Rating Scale

ABSTRACT

The Visual Analogue Scale (VAS), Numerical Rating Scale (NRS), Verbal Rating Scale (VRS), and the Faces Pain Scale-Revised (FPS-R) are among the most commonly used measures of pain intensity in clinical and research settings. Although evidence supports their validity as measures of pain intensity, few studies have compared them with respect to the critical validity criteria of responsivity, and no experiment has directly compared all 4 measures in the same study. The current study compared the relative validity of VAS, NRS, VRS, and FPS-R for detecting differences in painful stimulus intensity and differences between men and women in response to experimentally induced pain. One hundred twenty-seven subjects underwent four 20-second cold pressor trials with temperature order counterbalanced across 1°C, 3°C, 5°C, and 7°C and rated pain intensity using all 4 scales. Results showed statistically significant differences in pain intensity between temperatures for each scale, with lower temperatures resulting in higher pain intensity. The order of responsivity was as follows: NRS, VAS, VRS, and FPS-R. However, there were relatively small differences in the responsivity between scales. A statistically significant sex main effect was also found for the NRS, VRS, and FPS-R. The findings are consistent with previous studies supporting the validity of each scale. The most support emerged for the NRS as being both (1) most responsive and (2) able to detect sex differences in pain intensity. The results also provide support for the validity of the scales for use in Portuguese samples.



## 1. Introduction

The Visual Analogue Scale (VAS), Numerical Rating Scale (NRS), Verbal Rating Scale (VRS), and Faces Pain Scale-Revised (FPS-R) are among the most common measures of pain intensity used by clinicians and researchers. Evidence supports the reliability and validity of each of these measures across many populations [10,26–28]. However, each measure has strengths and weaknesses. For example, research indicates that VASs have more ratio scale qualities than other pain intensity scales for groups of patients (but not necessarily for individuals) [13,42,44,45], although some authors note that VASs scales do not always have linear qualities and are not always normally distributed [37,55]. Pain scales with more response levels (eg, the VAS or 0–10 NRS relative to the 6-point FPS-R or 4-point VRS) have the potential to be more sensitive [6,9,55], although more response categories do not necessarily translate to more responsivity [7,19,29]. Furthermore, research findings suggest that no single measure is consistently more responsive than any of the other measures [6,9–11,28,30,36], although the responsivity of the VAS, NRS, VRS, and FPS-R has yet to be directly compared in the same study. Also, the validity of these scales has never been examined in a sample of individuals from Portugal. Evaluations of common pain measures in samples from different countries and cultures can help establish the cross-cultural generalizability of validity findings.

Perhaps the most important validity criterion for a pain measure is its ability to detect changes in pain with pain treatment or procedures known to produce pain. One method for doing this would be to use an experimental design in which the amount of stimulation is highly controlled [14]. The cold-pressor test is an experimental method for inducing pain that is thought to reflect many (but not all) of the critical components of clinical pain [25], and its advantages are discussed in the literature [14,22,40,53]. Moreover, an increase in pain intensity as water temperature decreases is well documented, with small variations in water temperature resulting in significant differences in pain intensity [22,53,54].

A number of studies have examined the influence of sex on pain perceptions and pain response to experimental pain [14,40], with

* Corresponding author. Address: Rua 25 de Abril, n° 5, Idanha – Belas, Belas 2605-119, Portugal. Tel.: +351 969082988.
*E-mail address:* mafvalente@gmail.com (M.A. Ferreira-Valente).

0304-3959/$36.00 
doi:10.1016/j.pain.2011.07.005

normally menstruating women usually being more sensitive to painful stimuli than men [34,36,51,54]. Laboratory pain experiments that include both men and women provide an opportunity to compare the ability of pain ratings to detect these well-established sex differences.

The primary aim of the current study was to compare the relative validity of VAS, NRS, VRS, and FPS-R for detecting differences in painful stimulation and for detecting sex effects in response to painful stimulation. Based on previous research, we hypothesized that all 4 scales would be able to detect a sex effect and differences in pain resulting from 4 temperatures (1°C, 3°C, 5°C, and 7°C). Given the larger number of response levels of the VAS and NRS, we also anticipated that these would evidence greater responsivity than the VRS or FPS-R. Finally, this study also sought to evaluate the validity of the pain intensity rating scales in a Portuguese sample.

## 2. Methods

### 2.1. Participants

Participants were 127 volunteer university students. Exclusion criteria included: (1) being under 18 years of age; (2) reporting a history of any of the following diagnoses or medical problems: musculoskeletal problems, cancer, heart disease, stroke, epilepsy, diabetes or Raynaud syndrome; (3) having an open wound, cut, or fracture in any of the upper limbs; (4) having a cognitive or physical disability that could prevent participation; or (5) refusal to participate.

Of the 127 subjects who expressed an interest in participating in the study, 112 were eligible and completed the entire experimental procedures. Of the 112 completers, 3 subjects were excluded from the analyses because they were unable to understand how to use the VAS. Thus, complete data were available for 109 subjects, 56 of whom were female (51.4%). The ages of the participants ranged from 18 to 40 years old (M = 22.27, SD = 3.92; Female: M = 21.24, SD = 3.45; Male: M = 23.34, SD = 4.13). Most of the sample had their permanent residence in an urban area (74.3%), and the remainder (25.7%) lived in a rural area. Ninety-nine participants were undergraduate college students (33.0%, 17.4%, 21.1%, 11.0%, and 8.3% in their first, second, third, fourth, and fifth year, respectively). Ten (9.2%) of the participants were in graduate school.

### 2.2. Material

The cold-pressor apparatus used consisted of 4 thermal insulated containers with 18.1 L of capacity containing water chilled to 4 different temperatures. The apparatus was capable of maintaining water temperatures within ±0.5°C of the desired temperatures throughout the experimental procedures. Each container had 2 compartments separated by a metal filter, one of which held water, ice, and a water pump, and the other (the immersion tank) contained water alone with an armrest; the participants' hands therefore never came in direct contact with the ice. Four water pumps (JAD, model SP-602, 200 L/h, Guangdong, China) made the water flow continuously between the 2 compartments of each container to prevent warm water pockets from forming near the participants' hands [53]. The temperature of the hand and water was monitored and controlled by asking the participants to hold a mercury thermometer in the palm of their hands and immersing a thermometer in the water, respectively. One extra thermal insulated container without divisions contained an armrest and tepid water.

### 2.3. Measures

Fig. 1 presents all the pain intensity rating scales used in this study. The VAS [23] consists of a horizontal line 100 mm in length,

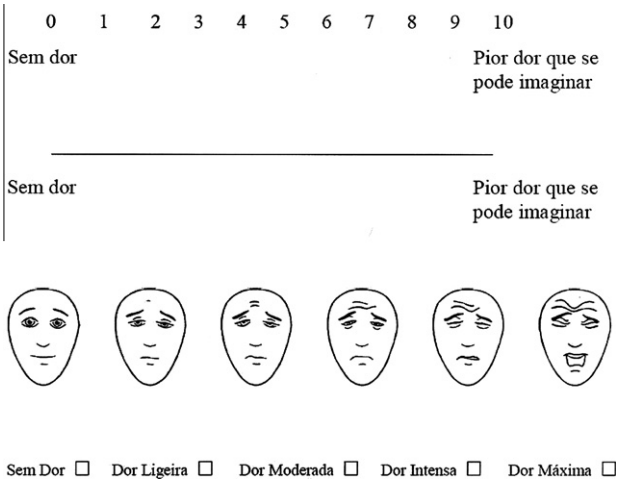


**Fig. 1.** The Numerical Rating Scale, Visual Analogue Scale, Faces Pain Scale – Revised, and Verbal Rating Scale; Faces Pain Scale – Revised, copyright ©2001, International Association for the Study of Pain, reproduced with permission, www.painsourcebook.ca.

with the end points "No pain" and "Worst imaginable pain" placed at each end of the line. Respondents are asked to make a mark on the line that best represents the level of pain intensity that they are experiencing. The NRS is an 11-point scale consisting of integers from 0 through 10; 0 representing "No pain" and 10 representing "Worst imaginable pain." Respondents select the single number that best represents their pain intensity. Although validity studies for the Portuguese versions of these measures have not been published, to our knowledge, both the VAS and the NRS used in this study have been previously used in research with Portuguese samples [1,16,17,48], and are recommended for use by the Portuguese Ministry of Health (Normative Circular n° 9/DGCG of June 14, 2003). The VRS is a 5-point scale consisting of a list of phrases (no pain, mild pain, moderate pain, intense pain, maximum pain) that describe increasing levels of pain intensity. Respondents select the single phrase that best characterizes their pain intensity. The VRS used in this study is commonly used by Portuguese researchers (eg, [12]). The FPS-R [4,21] is a 6-point scale, with 6 different faces that represent increasing levels of pain intensity. Respondents are asked to select the one expression that best characterizes his or her pain intensity, from the left-most face ("No pain"), to the right-most face ("Very much pain"). Each illustration corresponds to a numeric score (0, 2, 4, 6, 8, or 10). Research supports the validity of each of the pain measures used in this study as measures of pain intensity [24,27,32,39,44,46,47]. Although the FPS-R was initially developed for use with children, researchers also use the measure in samples of individuals with cognitive and communication impairment. The Portuguese (Portugal) translation of the FPS-R was performed by Batalha [2] and is available online (www.painsourcebook.ca).

### 2.4. Procedure

The cold-pressor procedures closely followed the guidelines for this task described in the literature, and were adapted to fit the study aims [40,52,53]. The study had Institutional Review Board approval. The study procedures were described to all potential participants and each was given a written consent form to read and sign before any measures were administered. After signing the consent form, participants completed a demographic and medical history questionnaire in order to identify potential medical conditions that would prevent participation. Participants who met any of the exclusion criteria were then excluded from participation.

The nondominant hand temperature was measured in all participants, followed by hand washing to the wrist. The nondominant hand was then immersed to the wrist in the container with tepid water (36°C ± 1°C) for 2 minutes, in order to reduce preexisting differences in hand temperature [53]. Hand temperature was again measured, and participants were instructed to immerse the hand to the wrist in the first cold water container for 20 seconds. Participants were also told they could take the hand from the cold water at any moment if it felt too uncomfortable to continue. For each of the 4 study conditions, only 3 participants took their hands out of the water before the established tolerance time (20 seconds). After the 20-second cold immersion, when we anticipated that pain would be at its most intense, we administered paper-and-pencil versions of the 4 pain measures (VAS, NRS, VRS, and FPS-R). The measures were presented in a random order using a Latin Square design. After a 3-minute break, the participants were asked to immerse his or her hand into the tepid warm water again for 2 minutes, and hand temperature was again assessed. Participants underwent 4 trials with 4 different water temperatures (1°C, 3°C, 5°C, and 7°C), in counterbalanced order, as proposed by Mitchell and colleagues [40]. Each participant experienced each water temperature only once, and provided ratings using each of the 4 scales for each temperature; thus, each participant provided 16 ratings. Participants were not given any information regarding the water temperature in each container.

#### 2.4.1. Data analysis

We first computed medians, means, and SDs for demographic and study variables for descriptive purposes. We next computed Pearson correlations between the VAS, NRS, FPS-R, and VRS, for descriptive purposes. In order to compare the ability of VAS, NRS, FPS-R, and VRS to detect differences in pain stimuli resulting from 4 different temperatures, as well as the hypothesized sex main effect on pain intensity ratings, we then performed 4 mixed-design repeated-measures analyses of variance (ANOVAs), with the pain intensity ratings as the dependent variables, and sex and temperature as the independent variables. Prior to these analyses, we evaluated test assumptions, namely normality and sphericity of the variance–covariance matrix, by analysing skewness (Sk) and kurtosis (Ku), with values of Sk and Ku lower than 1 indicating absence of severe violation of normality assumption and Mauchly test, respectively [5,18]. If a violation of the assumption of sphericity was found, we planned to use Huynh-Feldt epsilon to set the degrees of freedom [8,38]. In the event that a significant temperature effect was found, we planned to perform between-temperature comparisons using post hoc Fisher's least significant difference tests. Effect sizes were estimated using $\eta_p^2$, and, along with *P* values and F statistic magnitudes, were used to compare the hypothesized differences in responsivity of the 4 pain measures, with larger $\eta_p^2$ and F statistics, as well as smaller *P* values, indicating greater sensitivity [10,11,24,49]. Finally, we performed power analyses to determine the sample size required to obtain significant effects for each of the 4 measures, both for the omnibus ANOVA and for each planned temperature paired comparison. Alpha was set at 0.05 and power at 0.95 for these analyses. Statistical analyses were computed using software PASW Statistics 18 (v. 18, SPSS Inc. Chicago, IL, USA) and G*Power (v. 3.1) [15].

## 3. Results

Table 1 lists the descriptive statistics of the study variables, and Table 2 presents the correlation coefficients between the pain measures. As can be seen, the pain scales showed strong to very strong and statistically significant inter-scale correlations (*r*s ranging from 0.79 to 0.96) for all 4 water temperatures.

Pain ratings, as measured by VAS, NRS, VRS, and FPS-R, showed normal distributions for each of the 4 temperatures (Sk <1 and Ku <1). However, for each measure, we noted a violation of the assumption of sphericity [VAS: W = 0.70, X2 (5) = 38.28, $p < 0.001$; NRS: W = 0.59, X2 (5) = 55.53, $p < 0.001$; VRS: W = 0.82, X2 (5) = 20.54, $p < 0.01$; FPS-R: W = 0.81, X2 (5) = 21.74, $p < 0.01$]. We therefore used Huynh-Feldt epsilon to determine the degrees of freedom in the analyses [8,38].

As hypothesized, there were statistically significant temperature main effects for each of the 4 scales used in the study [VAS: $F_{Huynh\text{-}Feldt}(2.48, 265.24) = 85.74$; $p < 0.001$, $\eta_p^2 = 0.45$; $\pi = 1$; NRS: $F_{Huynh\text{-}Feldt}(2.28, 243.64) = 93.49$; $p < 0.001$, $\eta_p^2 = 0.47$; $\pi = 1$; VRS: $F_{Huynh\text{-}Feldt}(2.74, 287.48) = 76.36$; $p < 0.001$, $\eta_p^2 = 0.42$; $\pi = 1$; FPS-R: $F_{Huynh\text{-}Feldt}(2.70, 286.42) = 72.62$; $p < 0.001$, $\eta_p^2 = 0.32$; $\pi = 1$]. Moreover, all of the effect sizes associated with the temperature main effects in the omnibus ANOVAs for each measure were large. Effect sizes for the comparisons for each pair of temperature differences ranged from medium (0.17) to large (0.59), consistent with the differences between temperatures. These differences also follow the same pattern for each scale (Table 3 and Fig. 2).

Statistically significant sex main effects emerged for the NRS [$F(1, 107) = 4.40$; $p < 0.05$, $\eta_p^2 = 0.04$; $\pi = 0.55$], VRS [$F(1, 107) = 8.38$; $p < 0.01$, $\eta_p^2 = 0.07$; $\pi = 0.82$], and FPS-R [$F(1, 107) = 14.13$; $p < 0.001$, $\eta_p^2 = 0.12$; $\pi = 0.96$], and a nonsignificant trend emerged for the main effect of gender on pain ratings on VAS [$F(1, 107) = 3.49$; $p = 0.07$, $\eta_p^2 = 0.03$; $\pi = 0.46$], with women reporting higher pain ratings in every experimental condition.

Post hoc Fisher's least significant difference paired temperatures comparisons for each pain scale found statistically significant differences between all observations ($p < 0.001$). As would be expected, in every case, lower temperatures resulted in higher pain intensity ratings for all 4 scales.

The NRS and VAS evidenced slightly higher effect sizes (0.47 and 0.44, respectively) and higher F statistics (93.49 and 85.74, respectively) than VRS (0.42 effect size and F statistic of 76.36) and FPS-R (0.32 effect size and F statistic of 72.62). Power analyses based on these effect sizes indicated that the number of participants needed to be able to detect an overall difference between temperatures, as tested by an ANOVA, would be 5, 5, 5, and 7 for the NRS, VAS, VRS, and FPS-R, respectively. Power analyses to

**Table 1**
Means and SDs of the pain ratings for each temperature condition.

| | Overall Mean (SD) | Male Mean (SD) | Female Mean (SD) |
|---|---|---|---|
| *VAS* | | | |
| 7°C | 3.65 (2.35) | 3.23 (2.29) | 4.03 (2.36) |
| 5°C | 4.38 (2.23) | 3.90 (2.10) | 4.82 (2.28) |
| 3°C | 5.04 (2.24) | 4.67 (2.31) | 5.38 (2.14) |
| 1°C | 5.83 (2.17) | 5.58 (2.24) | 6.05 (2.12) |
| *NRS* | | | |
| 7°C | 3.87 (2.39) | 3.43 (2.38) | 4.29 (2.33) |
| 5°C | 4.64 (2.21) | 4.13 (2.17) | 5.12 (2.15) |
| 3°C | 5.32 (2.15) | 4.96 (2.24) | 5.66 (2.01) |
| 1°C | 6.18 (2.21) | 5.83 (2.29) | 6.52 (2.10) |
| *FPS-R* | | | |
| 7°C | 3.33 (2.51) | 2.52 (2.30) | 4.11 (2.48) |
| 5°C | 4.15 (2.32) | 3.32 (2.19) | 4.95 (2.21) |
| 3°C | 4.95 (2.55) | 4.08 (2.60) | 5.78 (2.23) |
| 1°C | 5.76 (2.46) | 5.17 (2.46) | 6.29 (2.35) |
| *VRS* | | | |
| 7°C | 2.49 (0.93) | 2.26 (0.88) | 2.67 (0.93) |
| 5°C | 2.79 (0.86) | 2.55 (0.82) | 3.00 (0.85) |
| 3°C | 3.08 (0.83) | 2.87 (0.81) | 3.30 (0.79) |
| 1°C | 3.50 (0.80) | 3.34 (0.85) | 3.63 (0.71) |

VAS, Visual Analogue Scale; NRS, Numerical Rating Scale; FPS-R, Faces Pain Scale-Revised; VRS, Verbal Rating Scale.

**Table 2**
Inter-scale correlation coefficients between the VAS, NRS, FPS-R, and VRS.

| | 7°C | | | 5°C | | | 3°C | | | 1°C | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VAS | NRS | FPS-R | VAS | NRS | FPS-R | VAS | NRS | FPS-R | VAS | NRS | FPS-R |
| VAS | – | – | – | – | – | – | – | – | – | – | – | – |
| NRS | 0.96 | – | – | 0.95 | – | – | 0.94 | – | – | 0.95 | – | – |
| FPS-R | 0.84 | 0.85 | – | 0.79 | 0.81 | – | 0.80 | 0.80 | – | 0.84 | 0.84 | – |
| VRS | 0.80 | 0.82 | 0.81 | 0.80 | 0.79 | 0.82 | 0.81 | 0.81 | 0.86 | 0.80 | 0.80 | 0.84 |

VAS, Visual Analogue Scale; NRS, Numerical Rating Scale; FPS-R, Faces Pain Scale-Revised; VRS, Verbal Rating Scale.

**Table 3**
Effect sizes (and number of participants required to obtain significant effects) for the VAS, NRS, FPS-R, and VRS.

| | Effect size (n) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VAS | | | NRS | | | FPS-R | | | VRS | | |
| | 5°C | 3°C | 1°C | 5°C | 3°C | 1°C | 5°C | 3°C | 1°C | 5°C | 3°C | 1°C |
| 7°C | 0.22 (14) | 0.47 (6) | 0.58 (5) | 0.25 (12) | 0.49 (6) | 0.59 (5) | 0.20 (16) | 0.44 (7) | 0.55 (5) | 0.17 (18) | 0.42 (7) | 0.58 (5) |
| 5°C | – | 0.20 (16) | 0.46 (7) | – | 0.22 (14) | 0.48 (6) | | 0.20 (16) | 0.45 (7) | – | 0.17 (19) | 0.29 (10) |
| 3°C | – | – | 0.30 (10) | – | – | 0.36 (8) | – | | 0.20 (15) | – | – | 0.29 (10) |

VAS, Visual Analogue Scale; NRS, Numerical Rating Scale; FPS-R, Faces Pain Scale-Revised; VRS, Verbal Rating Scale.

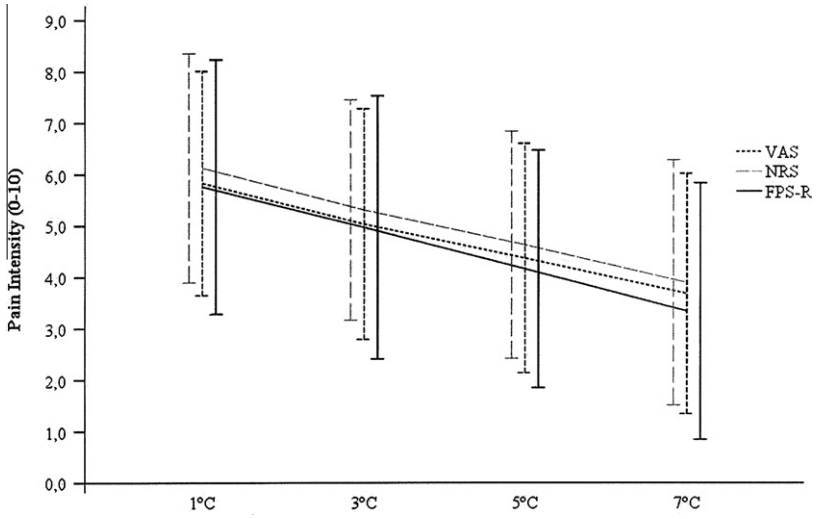


**Fig. 2.** Average pain intensity ratings across temperatures. Error bars represent SD.

compute the number of participants needed to detect differences between each pair of temperatures for each of the pain scales are presented in Table 3, and range from 5 to 19. These are consistent with the distance between temperatures, with larger distances (and effect sizes) corresponding to fewer participants needed.

## 4. Discussion

The results of this study provide strong support for the validity of all 4 scales studied for detecting changes in pain intensity in Portuguese university students. All of the scales were able to detect differences in pain resulting from 4 different temperatures, with variations in temperature resulting in statistically significant differences in pain intensity ratings, and lower temperatures resulting in higher pain ratings for each of the 4 scales studied. These results are consistent with previous studies that support each scale's validity [9,10,24,33,44,45], and show that small variations in water temperature result in significant differences in pain intensity ratings [22,40,53,54].

As predicted, we found some differences in the relative responsivity, with NRS being the most responsive, followed by VAS, VRS, and FPS-R. This is consistent with previous studies demonstrating the superiority of VAS and NRS responsiveness, due perhaps to the larger number of response levels that these 2 scales provide [6,9,41,50,55]. In our study, the NRS has shown to be slightly more responsive than the VAS, as indicated by its larger effect size and F statistic value. This is consistent with a group of studies showing a similar sensitivity between the NRS and VAS or a slight superiority of the NRS over the VAS [6,9–11]. Nevertheless, the results of power analyses to determine the number of subjects needed to detect a significant effect indicate that the 4 measures have very similar levels of responsivity, with very small differences between scales in the numbers of subjects needed to detect differences. This finding is consistent with the strong to very strong associations we found among the study measures (a finding also consistent with previous research, eg, [11,24,33,46]), indicating that all 4 measures tap into the same overall dimension (ie, pain intensity).

Thus, the results indicate that all else being equal, any of the 4 scales could be used for detecting changes in pain, although the NRS and VAS might be considered first when particularly sensitive and responsive measures of pain intensity are needed. Based on other considerations, however, researchers and clinicians may

elect to choose the NRS over the VAS in many settings. First, although the NRS has not consistently been shown to have ratio properties [43,45], its scores can provide data for parametric analysis [6,13,25,55]. Also, the NRS has been shown to be at least as sensitive as the VAS, whether a 0–10 NRS or a 0–100 NRS is used [6,9–11,24]. Third, the NRS is preferred over the VAS by patients and clinicians for its relative simplicity and ease of administration and scoring even when administered verbally [3,6,10,11,13,20,24,55]. Fourth, the VAS tends to have higher failure rates than the NRS or VRS, probably because both the NRS and the VRS are very easy to understand and complete by patients [3,6,10,11,13,20,24]. This latter point was also supported in the current study, given our finding that the only participants excluded from the sample for not being able to understand the measures were excluded because they were unable to understand the VAS.

The VRS is a categorical measure that might not have ratio properties [43]. As a result, VRSs do not necessarily have equal intervals between levels, which limits the conclusions that can be drawn about the magnitude of differences over time or between patient groups [25,26]. Also, when differences in responsivity are found, VRSs, as well as the FPS-R, tend to be less sensitive than VAS and NRS, consistent with the results from our study. This lower level of responsivity may be related to the lower number of response categories of these measures. However, we did find that both of these scales were able to detect changes in pain associated with differences in water temperature, indicating that they are valid and could be used when responsivity is not a critical issue [6,9,24,25,31]. Also, statisticians note that it is possible to draw valid conclusions using parametric analysis with ordinal data, such as data from VRSs, especially if number of categories in the scale is 5 or more [5,18,38]. Thus, the VRS can be considered a viable choice in settings with patients or research subjects who might be less able to use the NRS (eg, very young individuals or individuals with significant cognitive impairment). Likewise, our findings provide support for the validity of the FPS-R in adults, supporting its use in clinical and research settings. Although the FPS-R has been developed for use with children and proven useful with people with cognitive and communication disabilities, there may be situations or settings in which it might be useful for other populations. For example, the FPS-R might be useful for samples that include both adults without cognitive impairment *and* children (or the elderly) in the same study, and a measure is needed that all participants can complete. The FPS-R might also be considered for use in cross-cultural studies with adults, where researchers cannot be certain of the meaning equivalence of the verbal endpoint descriptors. In this situation, a measure based on facial expressions might show a greater cultural equivalence.

Regarding sex effect on pain intensity ratings, our results are in line with previous research showing significant sex effects on intensity ratings following painful stimulation [34,35,51,54], with women reporting higher pain intensity ratings across temperatures. The sex effects were statistically significant for 3 of the scales (NRS, VRS, FPS-R), and showed a nonsignificant trend for the VAS. It is interesting to note that other researchers have not found sex main effects when they used the VAS to test for these effects [40,45]. These findings suggest the possibility that for experiments specifically designed to test for or explain sex effects in pain intensity, it might be best to avoid using the VAS.

One significant limitation of the study is that it was performed with healthy young participants. Although it has been argued by some that the cold-pressor test mimics the effects of chronic pain conditions [40,53] due to its unpleasantness and to the fact that the painful stimuli is conducted by the C fibers, which are implicated in chronic pain, experimental pain is different from clinical pain. Clinical pain, for example, is less predictable and controllable. Also, participants in experimental pain studies can be assured that the pain is not associated with any tissue damage, whereas patients with clinical pain cannot always be so sure of this. For these reasons, clinical pain has an emotional significance and quality-of-life implications that may influence pain perception [14]. Therefore, the study findings do not necessarily generalize to patients with clinical pain conditions. It would be useful to examine the relative responsivity of the 4 pain measures in response to treatments or procedures known to impact clinical pain conditions to help determine their generalizability.

Nevertheless, the findings provide support for the validity and sensitivity of all 4 pain scales studied, with the exception that it might be best to avoid the use of VAS in studies seeking to examine sex effects. The findings also suggest that the NRS may be (very) slightly more sensitive than the other measures in our sample of individuals from Portugal; a finding consistent with some other studies that have compared the NRS to other pain measures, and supporting their cross-cultural reliability. Research is needed to compare these measures in clinical settings to confirm the generalizability of the current findings to clinical populations.

## Conflict of interest statement

None of the authors have any conflicts of interest with respect to this study.

## Acknowledgements

The authors gratefully acknowledge José Elísio Pereira for his contribution in the construction of the apparatus, and to Rita Ferreira and Vera Melo for their assistance in data collection. M. Alexandra Ferreira-Valente received PhD grant SFRH/BD/40956/2007 in the past year from the Portuguese Foundation for Science and Technology. José L. Pais Ribeiro received a sabbatical grant from FCT (SFRH/BSAB/982/2010) between January and April 2010. Mark P. Jensen received research support, consulting fees, or honoraria in the past year from Analgesic Research, Consultants in Behavioral Research, Endo, Fralex, Medtronic, Merck, Pfizer, Smith & Nephew, US Department of Education, US Department of Veterans Affairs, and the US National Institutes of Health.

## References

[1] Arantes S, Ferreira C, Lobo S, Moutinho R, Correia J, Carvalho J, Marcos A. Low back pain: the reality of our pain treatment unit. Dor 2007;15:22–7.
[2] Batalha L. Instructions for administering the Faces Pain Scale-Revised in languages other than English. Available from: <http://www.painsourcebook.ca> [accessed 20.06.11].
[3] Bergh I, Sjöström B, Odén A, Steen B. An application of pain rating scales in geriatric patients. Aging (Milano) 2000;12:380–7.
[4] Bieri D, Reeve R, Champion GD, Addicoat L, Ziegler J. The Faces Pain Scale for the self-assessment of the severity of pain experienced by children: development, initial validation and preliminary investigation for ratio scale properties. Pain 1990;41:139–50.
[5] Bollen KA. Structural equation with latest variables. New York: Wiley; 1989.
[6] Bolton JE, Wilkinson RC. Responsiveness of pain scales: a comparison of three pain intensity measures in chiropractic patients. J Manipulative Physiol Ther 1998;21:1–7.
[7] Bone M, Critchley P, Buggy DJ. Gabapentin in postamputation phantom limb pain: a randomized, double-bind, placebo-controlled, cross-over study. Pain 2002;27:481–6.
[8] Box GEP. Some theorems on quadratic form applied to the study of analysis of variance problems: II. Effects of inequality of variance and the correlation between errors in the two-way classification. Ann Math Stat 1954;25:484–98.
[9] Breivik EK, Bjornsson GA, Skovlund E. A comparison of pain rating scales by sampling from clinical trial data. Clin J Pain 2000;16:22–8.
[10] Bryce TN, Budh CN, Cardenas DD, Dijkers M, Felix ER, Finnerup NB, Kennedy P, Lundeberg T, Richards JS, Rintala DH, Siddall P, Widerstrom-Noga E. Pain after spinal cord injury: an evidence-based review for clinical practice and research. Report of the National Institute on Disability and Rehabilitation Research Spinal Cord Injury Measures meeting. J Spinal Cord Med 2007;30:421–40.
[11] Chanques G, Viel E, Constantin JM, Jung B, de Lattre S, Carr J, Cissé M, Lefrant JY, Jaber S. The measurement of pain in intensive care unit: comparison of 5 self-report intensity scales. Pain 2010;151:711–21.

[12] Cruz M, Gomes M, Sevivas N, Borralho N, Torres A, Silva B. Astragalus fracture – what's new – methods of diagnosis and treatment: 10 years experience. Rev Portuguesa Ortopedia Traumatol 2009;17:193–4 [Portuguese].
[13] Dijkers M. Comparing quantification of pain severity by verbal rating and numeric rating scales. J Spinal Cord Med 2010;33:232–42.
[14] Edens JL, Gil KM. Experimental induction of pain: utility in the study of clinical pain. Behav Ther 1995;26:197–216.
[15] Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39:175–91.
[16] Ferreira-Valente MA, Pais-Ribeiro JL, Jensen M. Coping with chronic muskuloskeletal pain: preliminary validation of the Portuguese version of two two item measures. Psychol Health 2009;24:171.
[17] Ferreira-Valente MA, Pais-Ribeiro JL, Jensen M. Adjustment and quality of life in persons with chronic musculoskeletal pain: importance of self-efficacy and social support. In: Livro das 18ª Jornadas da unidade de dor do Hospital Garcia de Orta; 2011. p. 30 [Portuguese].
[18] Finney SJ, DiStefano C. Non-normal and categorical data in structural equation modeling. In: Hancock GR, Mueller RO, editors. Structural equation modeling: a second course. Greenwich, CT: Information Age Publishing (IAP); 2006. p. 269–314.
[19] Frost S, Grossfeld S, Kirkley A, Litchfield B, Fowler P, Amendola A. The efficacy of femoral nerve block in pain reduction for outpatient hamstring anterior cruciate ligament reconstruction: a double-blind, prospective, randomized trial. Arthroscopy 2000;16:243–8.
[20] Hartrick CT, Kovan JP, Shapiro S. The numeric rating scale for clinical pain measurement: a ratio measure? Pain Pract 2003;3:310–6.
[21] Hicks CL, von Baeyer CL, Spafford P, van Korlaar I, Goodenough B. The Faces Pain Scale – Revised: toward a common metric in pediatric pain measurement. Pain 2001;93:173–83.
[22] Hirsch MS, Liebert RM. The physical and psychological experience of pain: the effects of labelling and cold pressor temperature on three pain measures in college women. Pain 1998;77:41–8.
[23] Huskisson E. Visual analogue scales. In: Melzack R, editor. Pain measurement and assessment. New York: Raven Press; 1983. p. 33–7.
[24] Jensen MP. The validity and reliability of pain measure in adults with cancer. J Pain 2003;4:2–21.
[25] Jensen MP. Pain assessment in clinical trials. In: Wittink H, Carr D, editors. Pain management: evidence, outcomes, and quality of life in pain treatment. Amsterdam: Elsevier; 2008. p. 57–88.
[26] Jensen MP. Measurement of pain. In: Fishman SM, Ballantyne JC, Rathmell JP, editors. Bonica's management of pain. Media, PA: Williams & Wilkins; 2010. p. 251–70.
[27] Jensen MP, Chen C, Brugger AM. The relative validity of three pain treatment outcome measures in post-surgical pain. Pain 2003;99:101–9.
[28] Jensen MP, Miller L, Fisher LD. Assessment of pain during medical procedures: a comparison of three scales. Clin J Pain 1998;14:343–9.
[29] Jensen MP, Turner JA, Romano J. What is the maximum number of levels needed in pain intensity measurement? Pain 1994;58:387–92.
[30] Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. Pain 1999;83:157–62.
[31] Joyce CR, Zutshi DW, Hrubes V, Mason RM. Comparison of fixed interval and visual analogue scales for rating chronic pain. Eur J Clin Pharmacol 1975;8:415–20.
[32] Kahl C, Cleland JA. Visual analogue scale, numeric pain rating scale and the McGill Pain Questionnaire: an overview of psychometric properties. Phys Ther Rev 2005;10:123–8.
[33] Kim EJ, Buschmann MT. Reliability and validity of the Faces Pain Scale with older adults. Int J Nurs Stud 2006;43:447–56.
[34] Klatzkin RR, Mechlin B, Girdler SS. Menstrual cycle phase does not influence gender differences in experimental pain sensitivity. Eur J Pain 2010;14:77–82.
[35] Kowalczyk WJ, Evans SM, Bisaga AM, Sullivan MA, Comer SD. Sex differences and hormonal influences on response to cold pressor pain in humans. J Pain 2006;7:151–60.
[36] Kucuk O, Fisher E, Moinpour CM, Coleman D, Hussain MH, Sartor AO, Chatta GS, Lowe BA, Eisenberger MA, Crawford ED. Phase II trial of bicalutamide in patients with advanced prostate cancer in whom conventional hormonal therapy failed: a Southwest Oncology Group study (SWOG 9235). Urology 2001;58:53–8.
[37] Lund I, Lundeberg T, Sandberg L, Budh CN, Kowalski J, Svensson E. Lack of interchangeability between visual analogue and verbal rating pain scales: a cross sectional description of pain etiology groups. BMC Med Res Methodol 2005;5:31.
[38] Maroco J. Análise estatística com utilização do SPSS (3ª Edição). Lisboa: Edições Sílabo; 2007 [Portuguese].
[39] Miró J, Huguet A. Evaluation of reliability, validity, and preference for a pediatric pain intensity scale: the Catalan version of the faces pain scale-revised. Pain 2004;111:59–64.
[40] Mitchell LA, MacDonald RA, Brodie EE. Temperature and the cold pressor test. J Pain 2004;5:233–8.
[41] Moore MJ, Osoba D, Murphy K, Tannock IF, Armitage A, Findlay B, Coppin C, Neville A, Venner P, Wilson J. Use of palliative end points to evaluate the effects of mitoxantrone and low-dose prednisone in patients with hormonally resistant prostate cancer. J Clin Oncol 1994;12:689–94.
[42] Myles PS, Urquhart N. The linearity of the visual analogue scale in patients with severe acute pain. Anaesth Intensive Care 2005;33:54–8.
[43] Price DD, Bush FM, Long S, Harkins SW. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. Pain 1994;56:217–26.
[44] Price D, McGrath P, Rafii A, Buckingham B. The validation of Visual Analogue Scales as ratio scale measure for chronic and experimental pain. Pain 1983;17:45–56.
[45] Price DD, Patel R, Robinson MR, Staud R. Characteristics of electronic visual analogue and numerical scales for ratings of experimental pain in healthy subjects and fibromyalgia patients. Pain 2008;140:158–66.
[46] Ramer L, Richardson JL, Cohen MZ, Bedney C, Danley KL, Judge EA. Multimeasure pain assessment in an ethnically diverse group of patients with cancer. J Transcult Nurs 1999;10:94–101.
[47] Shannon MM, Ryan MA, D'Agostino N, Brescia FJ. Assessment of pain in advanced cancer patients. J Pain Symptom Manage 1995;10:274–8.
[48] Silveira H, Soares J, Lima T, Sousa A, Brasil R, Soares C, Lima H, Carreiro P. Pain after amygdalectomy – classic dissection vs electro-dissection. Dor 2002;10:9–11 [Portuguese].
[49] Stinson JN, Kavanagh T, Yamada J, Gill N, Stevens B. Systematic review of the psychometric properties, interpretability and feasibility of self-report pain intensity measures for use in clinical trials in children and adolescents. Pain 2006;125:143–57.
[50] Stockler MR, Osoba D, Goodwin P, Corey P, Tannock IF. Responsiveness to change in health-related quality of life in a randomized clinical trial: a comparison of the Prostate Cancer Specific Quality of Life Instrument (PROSQOLI) with analogous scales from the EORTC QLQ-C30 and a trial specific module. European Organization for Research and Treatment of Cancer. J Clin Epidemiol 1998;51:137–45.
[51] Tashani OA, Alabas OA, Johnson MI. Cold pressor pain responses in healthy Libyans: effect of sex/gender, anxiety, and body size. Gend Med 2010;7:309–19.
[52] Trapanotto M, Pozziani G, Perissinotto E, Barbieri S, Zacchello F, Benini F. The cold pressor test for the pediatric population: refinement of procedures, development of norms, and study of psychological variables. J Pediatr Psychol 2009;34:749–59.
[53] von Baeyer CL, Piira T, Chambers CT, Trapanotto M, Zeltzer LK. Guidelines for the cold pressor task as an experimental pain stimulus for use with children. J Pain 2005;6:218–27.
[54] Walsh NE, Schoenfeld L, Ramamurthy S, Hoffman J. Normative model for cold pressor test. Am J Phys Med Rehabil 1989;68:6–11.
[55] Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. J Clin Nurs 2005;14:798–804.