Title page

Development and Evaluation of Linguistic Stimuli for Pain Research

Julie F Vermeir[1], BBehSc (Hons); Melanie J White[1], PhD; Daniel Johnson[2], PhD; Geert Crombez[3], PhD; Dimitri M L Van Ryckeghem[3,4,5], PhD

[1]Queensland University of Technology (QUT), Faculty of Health, School of Psychology and Counselling, Brisbane, Australia

[2]Queensland University of Technology (QUT), Faculty of Science, Brisbane, Australia

[3]Ghent University, Department of Experimental Clinical and Health Psychology, Ghent, Belgium

[4]Maastricht University, Department of Clinical Psychological Science, Maastricht, Netherlands

[5]University of Luxembourg, Department of Behavioural and Cognitive Sciences, Esch-sur-Alzette, Luxembourg

Corresponding author: Julie F. Vermeir. *Address*: Faculty of Health, School of Psychology and Counselling, Queensland University of Technology (QUT), 170 Victoria Park Road, Brisbane, QLD, 4059, Australia. *Phone*: +61 731384714. *Email address*: julie.vermeir@hdr.qut.edu.au

Running title: Linguistic stimulus material for pain research

Disclosures: The study described in the manuscript is part of a PhD for Julie F. Vermeir, which is funded by the Australian Government Research Training Program and Queensland University of Technology (QUT). Dimitri M. L. Van Ryckeghem is supported by the FNR Core Junior programme (Painflex: Nr. 12671141). All authors declare no conflicts of interest.

Abstract

Linguistic stimuli are commonly used in research to investigate the processing of pain. To provide researchers with a dataset of pain-related and non-pain-related linguistic stimuli, this research investigated 1) the associative strength between pain-related words and the pain construct; 2) the pain-relatedness ratings of pain words; and 3) the variability in the relatedness of pain words within pain word classifications (e.g., sensory pain words). In Study 1, 194 pain-related and matched non-pain-related words were retrieved by reviewing the pain-related attentional bias literature. In Study 2, adults with (n=85) and without (n=48) self-reported chronic pain completed a speeded word categorisation paradigm and rated the pain-relatedness of a subset of pain words. Analyses revealed that 1) despite differences in associative strength of 11.3% of the words between chronic and non-chronic pain groups, no overall group difference was found, 2) the chronic pain group rated the pain words as more pain-related compared to the non-chronic pain group, and 3) there was variability in the relatedness of pain words within pain word classifications. The findings highlight the importance of validating linguistic pain stimuli. The resulting dataset is openly accessible and new published sets can be added to the *Linguistic Materials for Pain (LMaP) Repository*.

Perspective: This article presents the development and preliminary evaluation of a large pool of pain-related and non-pain-related words in adults with and without self-reported chronic pain. Findings are discussed and guidelines are offered to select the most suitable stimuli for future research.

Keywords: pain, chronic pain, stimulus material, word stimuli, adults

Introduction

Researchers in the field of clinical and health psychology have often used linguistic stimuli as experimental material to investigate the processing of threatening information in a wide array of health conditions, including anxiety [2], depression [33], fatigue [22], and pain [50]. More particularly, studies in pain have largely employed pain-related words in paradigms to trigger a range of cognitive [39,44], physiological [52], and neurological [36] responses. Linguistic stimuli have the advantage of being relatively quick to process, easy to implement, and their physical characteristics (e.g., word length) can be tightly controlled [18]. However, one of the methodological challenges in conducting pain research is selecting a set of suitable stimuli for a project given accessible linguistic stimulus sets are currently largely lacking. To date, researchers have typically relied on words selected from previous published studies or used the pain descriptors from the McGill Pain Questionnaire [30] (MPQ) without further assessing whether these words are relevant for the target population and hoped that the stimuli do in fact elicit the intended response in study participants [4,5,10,23].

Stimuli that are strongly associated with the concern of pain are crucial as they may facilitate activation of personal pain schemata [11]. In many cognitive tasks, word stimuli are only presented for a short duration (<500 ms), whereby initial recognition and associative strength between the word and its respective category (i.e., pain) may be key to revealing biased processing. Evidence for category-item associations was obtained by Fazio et al [16] using a response latency measure. Particularly, they found that reaction times (RTs) were faster for strong associates than for weak associates of a category, indicating that concepts are linked with varying strength and the stronger the association, the quicker the item will be retrieved from memory. Applying this finding to the pain context, automatic processing of affect may only be observed for those stimuli that have a strong associative strength. To provide preliminary information about the items, we chose a response latency measure to assess the

associative strength between the word and its respective category as well as a Likert rating scale to evaluate the pain-relatedness of pain words. In addition, the items were evaluated in both individuals with and without chronic pain given research suggests that pain-related words may be processed differently among these populations [14,37,40].

In the present research, the literature of attentional bias (AB) was first reviewed to create a pool of linguistic items for evaluation (Study 1). Then, this pool of items was used in a speeded word categorisation paradigm and with self-report ratings to assess the associative strength with the pain construct and their relatedness for pain, respectively, in both individuals with and without chronic pain (Study 2). The first objective was to examine whether the associative strength between pain-related words and their respective category would be stronger (i.e., faster RTs) in the chronic pain group compared to the non-chronic pain group. The second objective was to investigate whether the pain-relatedness of pain words would be rated higher in the chronic pain group compared to the non-chronic pain group. The last objective was to explore whether sensory pain words were related more to pain (i.e., faster RT and higher self-report ratings) compared to affective pain words, ill-health words and threat words. In doing so, this study aimed to 1) investigate the characteristics of pain-related linguistic stimuli as well as 2) provide a large dataset of pre-evaluated linguistic stimuli containing both pain-related and matched non-pain-related words, which can assist researchers in selecting stimuli that might be more suitable for their research objectives and target population.

Study 1: Identification and selection of pain-related and non-pain-related stimulus material

Methods

Search strategy

To create a pool of items for evaluation, we first reviewed linguistic stimuli used in dot-probe studies investigating AB towards pain. This area of research was selected because most studies investigating AB for pain have used both pain-related and non-pain-related (neutral) words as experimental material across a range of pain and pain-free populations [44], thereby providing a large pool of linguistic material for evaluation. Electronic searches of Scopus and Web of Science databases were conducted on 11 June 2018 using the following full and truncated search terms: (dot-probe OR visual-probe) AND word* AND (pain OR pain-related) AND "attention* bias*". No limiters were applied at this stage. A complementary manual search of the reference lists of key papers was also conducted to locate records not identified in the database searches.

Eligibility criteria

Studies were selected if they met the following inclusion criteria:

1. Original empirical research that investigated AB (i.e., the tendency to prioritise attentional processing of pain-related information) within the context of pain.
2. Peer-reviewed document.
3. Employed a word-based dot-probe paradigm [27] to measure AB or used a modified version of the paradigm [28] to manipulate AB.
4. Used English words that authors attribute to pain. This includes sensory, affective, ill-health and threat pain word classifications.

5. Provided a partial or complete list of pain-related words within the article or in an online appendix. This does not include examples of words. Study authors who did not include a list of words were not contacted or included in the review; however, when reference was made to selecting words from previous research, that research was located and included in the review if the inclusion criteria were met.

Data extraction

The first author (JV) extracted pain-related and non-pain-related words into a Microsoft Excel spreadsheet. We only retained non-pain-related stimuli that were paired with pain-related stimuli (e.g., non-pain-related filler words were not considered).

Results

Study selection

Figure 1 presents the PRISMA flowchart for the selection of studies included in this review. The database searches identified 84 records and searches through other sources retrieved an additional 22 records. The first author (JV) identified and removed duplicates using Endnote X8 citation management. Next, records were screened by title and abstract (n=74), and studies clearly situated outside the topic area (e.g., assessing food-related AB) were excluded. The remaining records (n=56) were then assessed for eligibility based on their full-text. A total of 20 records, reporting on 21 independent studies were included in the review. The key characteristics of the 21 studies included in the review are summarised in Multimedia Appendix 1.

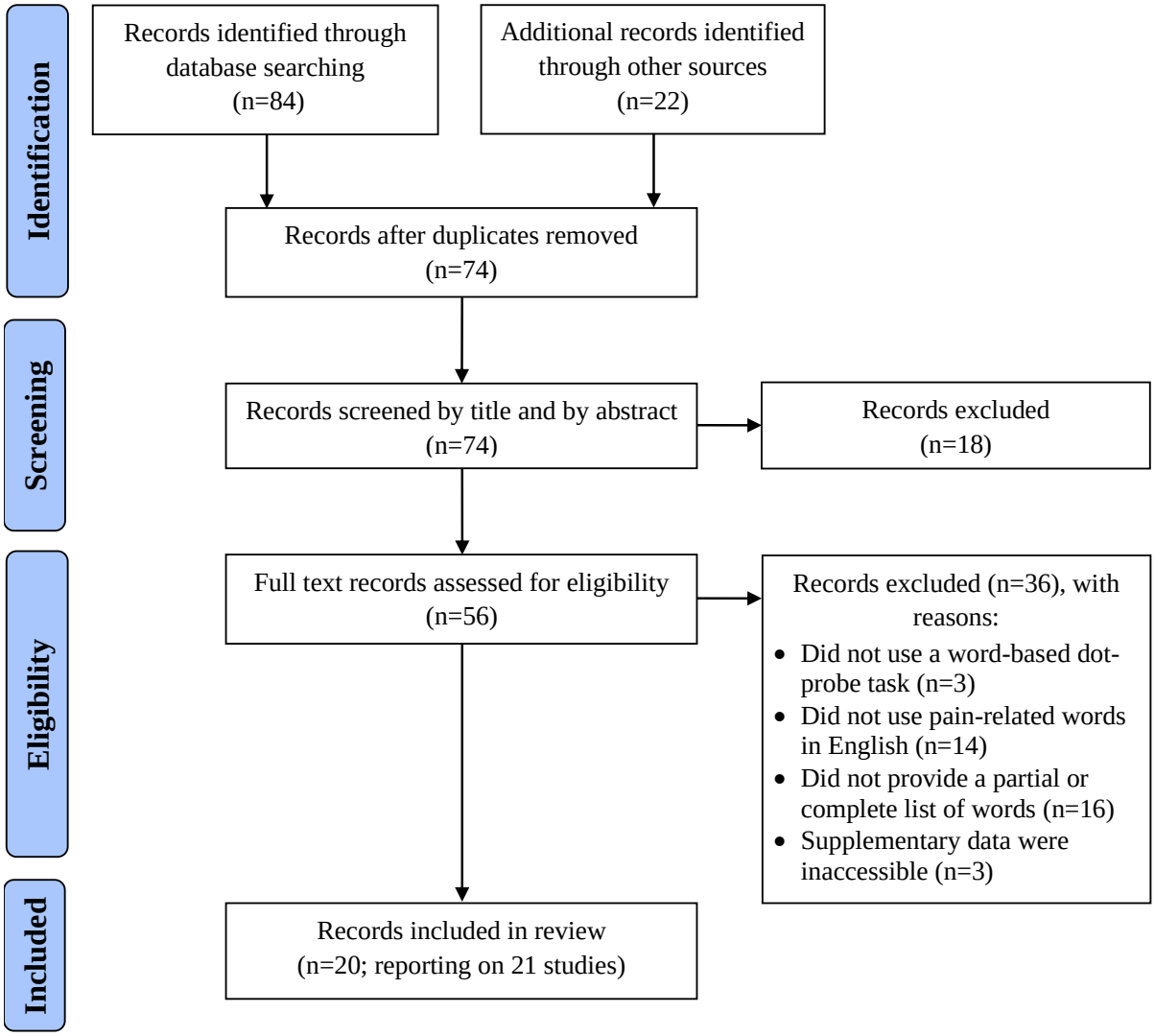

Figure 1. PRISMA flowchart of the study selection process.

Identification of pain-related and non-pain-related words

The review identified 194 different pain-related words and 336 matched non-pain-related words (duplicates removed). The most widely used pain stimulus was the word *throbbing* (n=16/21; 76.2%), followed by the words *shooting* (n=15/21; 71.4%) and *sharp* (n=15/21; 71.4%). Interestingly, few studies used the word *pain* (n=7/21; 33.3%) and *painful* (n=4/21; 19.0%), despite being basic pain terms. It is unknown why these pain descriptors have not been used more frequently as this was not explicitly stated in the reviewed studies. It is possible that these pain descriptors were not used for reasons of transparency and priming. Regarding non-pain-related stimuli, most words were drawn from the category of household

items (e.g., *saucepan*). A complete list of pain-related and matching non-pain-related words for each of the included studies can be found in Multimedia Appendix 1.

Selection of pain-related and non-pain-related words for evaluation

A total of 194 different pain-related words were identified in the review and chosen for the evaluation procedure. For non-pain-related stimulus material, we only retained 193 of the 336 words collected, due to feasibility reasons. The stimuli were selected by JV and DVR in the following ways: 1) when more than one non-pain-related word was paired to a pain word, the non-pain-related word that best matched the pain word in terms of number of letters and written frequency (as reported by the original authors) was retained; 2) when the same non-pain-related word was paired to more than one pain word, another non-pain-related word within the list that best matched the pain word in terms of number of letters (as a minimum) and written frequency (when possible) was selected; and 3) when a non-pain-related word was not matched on length to a pain word either deliberately (i.e., authors reported that not all word pairs were matched on length) or by error (i.e., authors reported that all word pairs were matched on length; however, this was not always the case), one within the list that met that criteria was selected, except for the word *delegation* which was made plural. Finally, for the pain-related word *incomprehensible* there was no matched non-pain-related word nor a word in the list that matched on length; therefore, we self-selected from beyond this dataset the non-pain-related word *intercommunication* based upon the number of letters. In summary, a total of 194 pain-related words and 194 unique non-pain-related words were chosen for the evaluation tasks.

Study 2: Evaluation of pain-related and non-pain-related stimulus material

Methods

Participants

Participants were recruited between January 2019 and April 2020, from the university and wider Australian community, through distribution of flyers, emails sent to student and staff mailing lists, social media, and other channels (e.g., Diabetes Australia). Inclusion criteria were: 1) being aged 18 or older; 2) being a native English speaker; 3) having access to a desktop or laptop computer connected to reliable internet (to avoid potential confounding influences such as substantial differences in screen sizes and keyboards resulting from the use of tablets and smartphones); and 4) either experiencing self-reported chronic neuropathic pain (i.e., pain arising from a lesion or disease of the nervous system lasting for longer than 3 months), chronic musculoskeletal and/or fibromyalgia pain (i.e., pain arising from the bones, joints, muscles, or related soft tissues lasting for more than 3 months), or having no history of chronic pain. First-year psychology students were given course credit for their participation and all other participants received an online shopping voucher of AUD$10. Power analyses (G-power [15]) indicated that a minimum of 159 participants (53 participants per group) were needed to have sufficient power (0.8) to detect a medium effect size. Therefore, we aimed to recruit 168 participants (56 participants per group) to account for potential drop-out due to careless responding. Data collection ended when numbers had been reached, except for the neuropathic pain group where the quota was not attained due to recruitment difficulties.

A diagram depicting participant flow through the study is presented in Figure 2. A total of 683 volunteers agreed to participate in this study. Of these, 53 participants were excluded as they did not meet the inclusion criteria. Consequently, 630 participants started the survey, but

only 255 completed it and began the evaluation tasks (40.5% participation rate). A further 104 participants were excluded due to dropping out at or shortly after launch (n=74), using a tablet or smartphone (see exclusion criteria; n=16), technical difficulties which led to no data being recorded (n=5), and failing to respond correctly on the instructional question (n=9). An additional two participants from the non-chronic pain group were removed from the analyses for scoring a pain grade of 2 or 3 on the Graded Chronic Pain Scale (GCPS) [51], indicating that they might be experiencing moderate to severe chronic pain. Finally, after examination of the data from the evaluation tasks, 15 participants were excluded for labelling the word *pain* (control word) as not related to pain (speeded word categorisation paradigm), suggesting that they did not understand the task instructions or were not engaged with the task, and one participant was excluded for rating more than 30.0% of the questions faster than 500 ms (explicit pain rating task), indicating that the participant was merely *clicking through* the questions. This left a final total sample of 133 participants.

Although three groups of participants were recruited separately (i.e., chronic neuropathic pain, chronic musculoskeletal and/or fibromyalgia pain, and non-chronic pain), it was decided to combine the two chronic pain groups into a single chronic pain group because sample sizes were not sufficiently large to comprehensively examine the groups separately. The planned sample projections of minimum 56 participants per group could not be reached, particularly for the neuropathic pain group (n=36), despite 15 months of recruitment (January 2019 to April 2020) and intense advertising. Moreover, data from the two screening pain questionnaires revealed that nearly half (44.4%; n=16/36) of the participants in the neuropathic group met the diagnostic criteria for fibromyalgia and over one quarter (28.6%; n=14/49) of the participants in the musculoskeletal and/or fibromyalgia group screened for neuropathic pain, showing large overlap in symptoms in the groups that were recruited. Therefore, it was difficult to partition participants into two clearly defined chronic pain groups and excluding

from analyses participants who had overlapping conditions would have further reduced the sample size. Therefore, we considered it most prudent to analyse the pooled data, which increased the power of our analyses. The final sample comprised 85 adults with chronic pain and 48 adults without chronic pain.

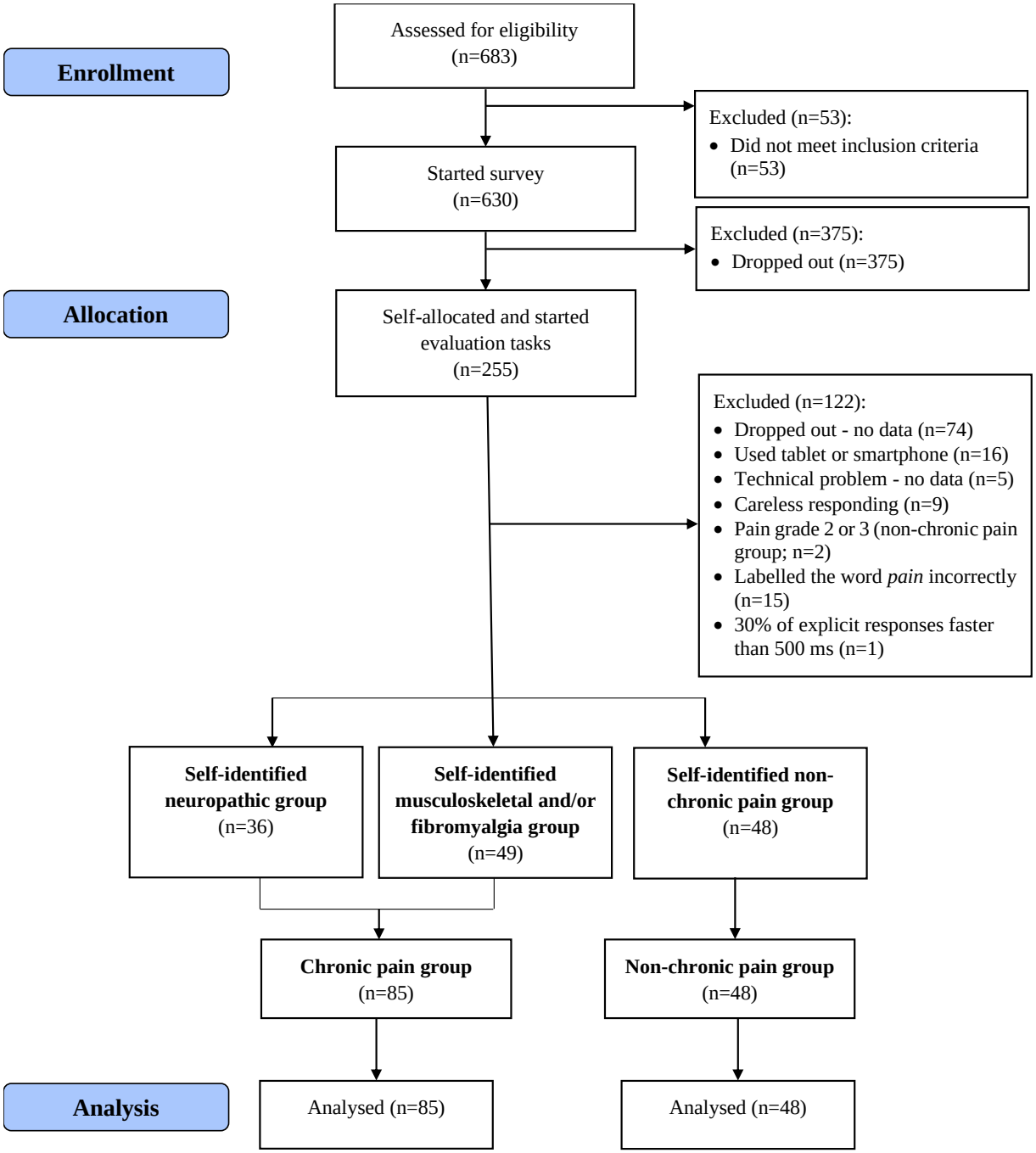

Figure 2. Study flow diagram. Note. The two self-identified chronic pain groups were heterogeneous with a large overlap in symptoms; therefore, the two pain groups were combined into a single chronic pain group.

Questionnaires

Demographic information

All participants reported on demographic information (e.g., age, gender, and country of birth) and health status (e.g., current general and mental health status). In addition, participants in the pain groups provided information on their pain experience (e.g., duration of primary pain condition).

Pain and affect questionnaires

The 7-item GCPS [51] is a self-report instrument designed to assess two dimensions of chronic pain severity (pain intensity and pain-related disability) in the general population and in primary health care settings. The scale measures the presence of chronic pain in the past 6 months and all items are scored on an 11-point Likert-scale, with responses ranging from 0 to 10. Subscale scores (i.e., characteristic pain intensity, disability score, and disability points) for the two dimensions are combined to calculate a chronic pain grade that allows individuals with chronic pain to be classified into 1 of 5 hierarchical categories: grades 0 (*no pain problem*) to 4 (*high disability – high intensity*). The GCPS has been found to have acceptable to excellent internal consistency, with a Cronbach α ranging from .74 to .91 [42,51].

The Neuropathic Pain Questionnaire [24] is a 12-item self-report questionnaire used to assess neuropathic pain and to discriminate between neuropathic and non-neuropathic pain. From these, 10 items are related to sensations or sensory responses and 2 items are about changes in sensitivity. Participants rate each item on a scale from 0 (*no pain*) to 100 (*worst imaginable pain/greatest increase*). After each item score is multiplied by a discriminant function coefficient, the scores are summed and incorporated with a set constant value (-1.408) to create a total discriminant function score. A discriminant function score below 0 predicts non-neuropathic pain, whereas a score at or above 0 predicts neuropathic pain. The

questionnaire has a sensitivity of 67.0% and a specificity of 74.0% compared to clinical diagnosis.

The Widespread Pain Index (WPI) and Symptom Severity Scale SS-Scale; [55] is a self-report fibromyalgia screening tool. The WPI assesses the presence of pain or tenderness over the past 7 days in 19 non-articular pain areas (e.g., left upper arm). Each item is scored as 0 or 1, with scores ranging from 0 to 19. The six-item SS-Scale consists of 2 parts. The first part asks participants to rate on a scale from 0 (*no problem*) to 3 (*severe, continuous, life-disturbing problems*) the symptoms of fatigue, trouble thinking or remembering, and unrefreshing sleep over the past 7 days. The second part asks participants to indicate if in the past 6 months they had any symptoms of "pain or cramps in lower abdomen", "depression", or "headache". Each symptom can be coded as "yes" or "no", scored as 1 or 0, respectively. The SS-Scale score can range from 0 to 12. It is considered that a person meets the diagnostic criteria for fibromyalgia if a WPI score is ≥7/19 and an SS-Scale score is ≥5/12 or if a WPI score is 3–6/19 and a SS-Scale score if ≥9/12 [54]. The diagnostic criteria for fibromyalgia have showed sensitivity and specificity values of 90.2% and 89.5%, respectively [17].

Negative affect was assessed with two Patient-Reported Outcomes Measurement Information System (PROMIS) measures comprising PROMIS Anxiety 8a (version 1.0; 8 items) and PROMIS Depression 8b (version 1.0; 8 items). The items have a 7-day period and are rated on a 5-point Likert scale ranging from 1 (*never*) to 5 (*always*). The raw score totals on each scale are transformed to T-score metrics using the PROMIS conversion tables, such that the average score for the general population is 50 and the standard deviation (*SD*)10 [21]. In line with previous research, T-scores <55 would translate as normal limits, 55-59 as mild, 60-70 as moderate, and ≥ 70 as severe emotional distress [8,25,53]. The two PROMIS measures have demonstrated excellent psychometric properties in both population-based [3] and clinical samples [1,38].

Instructional question

To identify careless responding patterns [29], one item from the Instructional Manipulation Check (i.e., Please select the option "Always") was added in the survey flow [32]. Participants who failed to answer the Instructional Manipulation Check correctly were excluded from further analyses.

Evaluation tasks

Word stimuli

A total of 194 different pain-related and 194 matched non-pain-related words were included in the evaluation procedure (see *Selection of pain-related and non-pain-related words for evaluation* section for further details). Based upon the coding system provided by Todd et al [44], pain-related words were classified as the following: 1) sensory pain words (describe the sensory qualities of pain); 2) affective pain words (describe the affective component of pain); 3) ill-health words (describe illness, injury, ill-health, general health threat); and 4) threat words (describe general threat, social threat, anger). Data was coded independently by three authors (JV, DVR, and MW). The Krippendorff's α test was used to estimate the inter-coder reliability [20], and results showed a high inter-coder reliability (α=.88). Following this, a total of 31 (16%) words needed arbitration. Any disagreement or ambiguity was resolved by discussion and consensus.

The pain dataset contains 86 (44.3%) sensory pain words (e.g., *hurting*), 42 (21.6%) affective pain words (e.g., *punishing*), 43 (22.2%) ill-health words (e.g., *sick*), and 23 (11.9%) threat words (e.g., *fearful*). The non-pain-related dataset contains a large variety of words, such as household items (e.g., *saucepan*), vehicles (e.g., *car*) and animals (e.g., *goat*). A complete list of pain-related and matching non-pain-related words is presented in Multimedia Appendix 1.

Speeded word categorisation paradigm

During the speeded word categorisation paradigm category-relevant labels (i.e., "PAIN" and "NO PAIN") are presented in the lower-left and lower-right corner of the screen (counterbalanced between participants). These labels remain present for the duration of the entire experiment, which consist of 396 trials (i.e., 388 critical trials containing 194 pain-related and 194 non-pain-related words, and 8 digit trials) divided across four blocks, with a pause offered between each block of trials. Target stimuli were presented in a randomised order across trials, whereby each word was presented once. All stimuli were presented in white 28-point uppercase Arial font on a black background.

To familiarise participants, the task started with five practice words selected from a neutral theme (i.e., earth-related words), with the presentation (right and left) of category-relevant labels counterbalanced across participants. Critical trials began with a 500 ms presentation of a white fixation cross (X) in the middle of the screen to direct attention to the centre of the screen (see Figure 3). Next, a target word was presented in the centre of the screen. The target word could be pain-related or non-pain-related, and the participants' task was to categorise as fast as possible whether the target stimulus was pain-related or not by pressing either the Q or P key on their keyboard, with the left and right index-finger, respectively. The target stimulus remained on the screen until response or 6000 ms after trial onset. In case participants answered too late, the term "too slow" appeared in the centre of the screen for 200 ms. In digit trials (n=8), the fixation cross was followed by a random digit number between 1 and 9 in the centre of the screen for a duration of 150 ms. Here, participants were required to indicate the digit using the keyboard. Digit trials were included to ensure that participants' attention was directed to the centre of the screen at the start of each trial [49]. The inter-trial interval was 1500 ms.

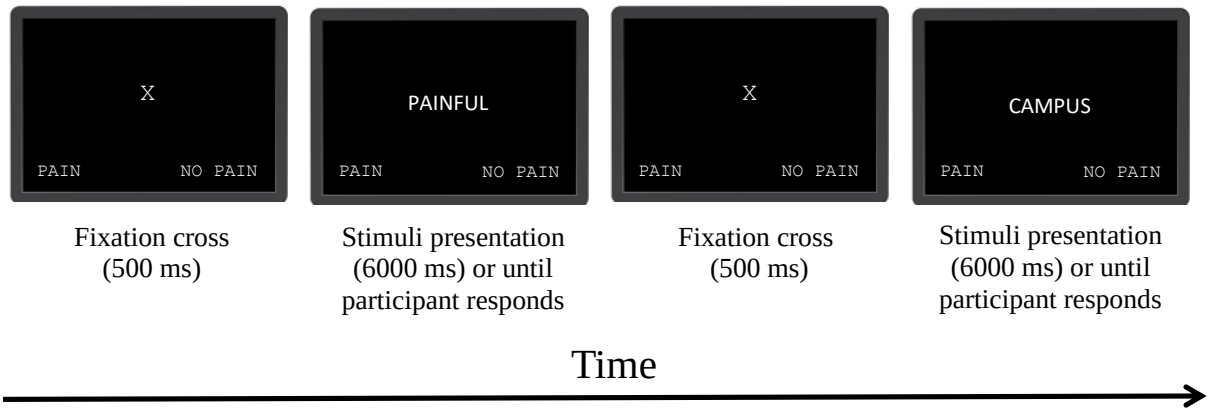

Figure 3. Time course of two sample trials from the speeded word categorisation paradigm. Stimuli are not presented to scale.

Explicit pain word ratings

To assess the relatedness of pain words, participants completed an explicit pain rating task. Here, we investigated the relatedness of the words with the construct "pain" rather than the relatedness with participants' pain condition only, although the general relatedness may not be independent from the chronic pain participants' own pain experience. This approach allowed us to use the same methodology for individuals not experiencing chronic pain at the time of testing and for those experiencing chronic pain.

In the explicit pain rating task, participants were presented with pain-related words in the centre of the screen and asked to rate how much the word is related to pain on a 11-point Likert scale, ranging from 0 (*not related*) to 10 (*highly related*). No time limit was imposed to rate the pain-relatedness of each word. To prevent participant fatigue and drop-out during the task, the 194 pain-related words were randomly assigned to four lists containing 48-49 words each, whereby participants were only presented one of the four word lists. The presentation order of the words within the list was completely randomised.

Procedure

The study was approved by Queensland University of Technology Human Research Ethics Committee (1800001005) and was conducted online at participants' time and place of convenience (using their own computers). Interested participants first provided informed

consent electronically, before being taken to the screening questions, and then to the survey, which was constructed using Key Survey (WorldAPP). All participants (i.e., both chronic pain and non-chronic pain groups) gave demographic information, completed the GPCS, the PROMIS Depression and Anxiety instruments, and answered a question to detect careless responding. In addition, participants with chronic pain provided information about their experience of pain and completed the Neuropathic Pain Questionnaire, WPI and SS-Scale. Next, participants in each group completed the speeded word categorisation paradigm and rated one of the four item sets for pain-relatedness, using Inquisit 5.0 (Inquisit Web Millisecond software package), which is a tool that allows the use of RT tasks in an online environment (https://www.millisecond.com/products/web). The study took approximately 45 minutes to complete.

Data treatment and statistical methods

Statistical analyses were conducted using the SPSS version 27.0 (IBM Corp). First, a series of analyses were performed to determine whether groups differed on key demographic variables, handedness, and affect measures. Next, we calculated for each word, the proportion of pain categorisation responses, the mean RTs separately for pain and no-pain categories (speeded word categorisation paradigm) as well as mean ratings for pain relatedness (explicit pain rating task). To compute proportion and mean RTs of categorisation responses, practice trials and all RTs below 150 ms (anticipations) were excluded [31,46,47]. The level of statistical significance was set at a *P* value <.05 (two-tailed) for all reported analyses. Where possible, effect sizes were calculated and presented by the test's most appropriate effect size [45].

For the individual word analyses, differences between the two groups (chronic pain vs. non-chronic pain) were analysed using chi-square or Fisher's exact tests for categorical variables (i.e., proportion of categorisation responses) and independent sample *t*-tests for continuous variables (i.e., mean ratings for pain-relatedness) or Mann-Whitney *U* tests, if the

data were not distributed normally ($P<.05$). To assess potential differences in mean RTs of categorisation responses across conditions, we used generalised linear models with gamma error distributions (i.e., right-skewed with a long tail in the slow RTs) and log link functions as it can account for the positive-skewed shape of the RT distribution without the need to transform and standardise the raw data [26]. The dependent variable was the amount of time participants spent on deciding whether the word displayed was pain-related or not.

In addition, a series of linear mixed model analyses were conducted. First, we investigated whether there were between groups (chronic pain vs. non-chronic pain) differences in participants' RTs and explicit pain ratings. Separate models for RT and explicit pain ratings were run. The non-chronic pain group served as the reference group for comparisons between groups and education status was entered as a covariate. Then, we investigated whether there were differences between groups (chronic pain vs. non-chronic pain) in participants' RTs and explicit pain ratings across the pain word classifications (i.e., sensory pain, affective pain, ill-health, and threat). Separate models for RT and explicit pain ratings were run. The non-chronic pain group served as the reference group for comparisons between groups, the sensory pain classification served as the reference point for comparisons between pain word classifications, and education level was entered as a covariate. We used a building procedure, with backward elimination (i.e., we removed one term at a time). The best fit model was identified by comparing models with a log likelihood ratio test and using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). All models incorporated a random intercept for participants and used the maximum likelihood estimator. RT and explicit pain rating data were non-normally distributed, so analyses were bootstrapped (percentile) using 1000 samples to provide robust probability values and confidence intervals.

Results

Participants characteristics

The final analysed sample contained 133 participants, aged 18 to 75 years (M_{age}=37.41, SD=16.17). Age was normally distributed, with skewness of 0.50 (SE=0.21) and kurtosis of -1.07 (SE=0.42). Of these, 84.2% (n=112/133) were female, 81.2% (n=108/133) were born in Australia, 78.9% (n=105/133) completed tertiary education (i.e., university, college, or post-high school qualifications), 45.9% (n=61/133) were single, 42.9% (n=57/133) were studying, and 88.7% (n=118/133) were right-handed.

Table 1 presents the demographic and descriptive statistics of key variables as a function of group. Overall, the chronic pain and non-chronic pain groups did not differ significantly on key demographic variables (i.e., age, gender, country of birth, cultural identity, relationship status, employment status, or handedness), except for education level, which was significantly lower in the chronic pain group ($X^2_{(1)}$=5.11, *P*=.024). As such, education level was added as a covariate to the linear mixed models. Compared to non-chronic pain subjects, participants with chronic pain also self-reported significantly worse general health (Fisher-Freeman-Halton's exact test=42.88, *P*=<.001) and more mental health problems ($X^2_{(1)}$=15.06, *P*=<.001), with PROMIS T-scores 0.5 SD higher than the standard population mean, indicating mild anxiety ($t_{(131)}$=-3.95, *P*=<.001) and depressive symptoms (*U*=1282.00, *P*=<.001). GCPS scores did differ significantly between groups (Fisher-Freeman-Halton's exact test=105.73, *P*=<.001), with all non-chronic pain subjects classified in the lower pain grades (i.e., Grade 0 or 1) and all participants with chronic pain classified in Grade 1 or above. Pain intensity was significantly ($t_{(131)}$=-11.05, *P*=<.001) higher in the chronic pain group (M=4.19; SD=2.22) compared to the non-chronic pain group (M=0.52; SD=0.80). Among the chronic pain group, 27 (31.8%) participants met the classification for neuropathic pain and 42 (49.4%) for fibromyalgia. The mean duration of pain was 6.94 years (SD=6.83 years; range: 3 months to

35 years). Over half of the participants self-reported having musculoskeletal or fibromyalgia pain (n=49/85; 57.6%) whereas the remaining participants self-reported having neuropathic pain (n=38/85; 42.6%).

Table 1

Key Demographic and Descriptive Statistics for Chronic and Non-chronic Pain Groups

Variable	Group		*P* value
	Chronic pain (n=85)	Non-chronic pain (n=48)	
Demographics			
Age, years [M (SD)]	37.69 (17.17)	36.92 (14.40)	[a].71
Gender, female [n (%)]	71 (83.5%)	41 (85.4%)	[b].82
Country of birth, Australia [n (%)]	70 (82.4%)	38 (79.2%)	[b].49
Handedness, right-handed [n (%)]	75 (88.2%)	43 (89.6%)	[b]1.00
Relationship status [n (%)]			[b].44
Single	38 (44.7%)	23 (47.9%)	
Married	22 (25.9%)	16 (33.3%)	
De facto relationship	10 (11.7%)	7 (14.6%)	
Divorced or separated	11 (12.9%)	2 (4.2%)	
Widowed	2 (2.4%)	0 (0.0%)	
Other	2 (2.4%)	0 (0.0%)	
Education level [n (%)]			[c].024*
Secondary (Year 10 or 12 certificate)	23 (27.1%)	5 (10.4%)	
Tertiary (university, college, or post-high school qualifications)	62 (72.9%)	43 (89.6%)	
Employment status [n (%)]			[b].27
Employed	30 (35.3%)	21 (43.8%)	
Student	34 (40.0%)	23 (47.9%)	
Retired	13 (15.3%)	3 (6.2%)	
Home duties	2 (2.4%)	0 (0.0%)	
Unemployed or unable to work	6 (7.0%)	1 (2.1%)	
General health status [n (%)]			[b]<.001*
Very good	5 (5.9%)	23 (47.9%)	
Good	31 (36.5%)	19 (39.6%)	
Fair	36 (42.3%)	6 (12.5%)	
Bad	13 (15.3%)	0 (0.0%)	
Psychological variables [M (SD)]			
Mental health condition, yes [n (%)]	41 (48.2%)	7 (14.6%)	[c]<.001*
PROMIS- Anxiety	58.47 (9.17)**	51.90 (9.29)	[d]<.001*
PROMIS- Depression	55.85 (10.35)**	49.78 (9.09)	[a]<.001*

Variable	Group		*P* value
	Chronic pain (n=85)	Non-chronic pain (n=48)	
GCPS pain intensity [M (SD)]	4.19 (2.22)	0.52 (0.80)	[d]<.001*
GCPS Scales [n (%)]			[b]<.001*
Grade 0: no pain problem	0 (0.0%)	11 (22.9%)	
Grade 1: low disability-low intensity	12 (14.1%)	37 (77.1%)	
Grade 2: low disability-high intensity	26 (30.6%)	0 (0.0%)	
Grade 3: high disability-low intensity	23 (27.1%)	0 (0.0%)	
Grade 4: high disability-high intensity	24 (28.2%)	0 (0.0%)	
NPQ-classification [n (%)]			
Neuropathic pain	27 (31.8%)	-	-
Non-neuropathic pain	58 (68.2%)	-	-
WPI and SS-Scale- classification [n (%)]			
Fibromyalgia	42 (49.4%)	-	-
No fibromyalgia	43 (50.6%)	-	-
Pain characteristics			
Duration of chronic pain, years [M (SD)]	6.94 (6.83)	-	-
Primary type of self-identified chronic pain [n (%)]			
Musculoskeletal and/or fibromyalgia pain	49 (57.6 %)	-	-
Neuropathic pain	36 (42.4%)	-	-

Note. SD=standard deviation; PROMIS=Patient-Reported Outcomes Measurement Information System; GCPS: Graded Chronic Pain Scale; NPQ=Neuropathic Pain Questionnaire; WPI and SS-Scale=Widespread Pain Index and Symptom Severity Scale.

[a] Mann-Whitney *U* test; [b] Fisher-Freeman-Halton's exact test; [c] chi-square test; [d] independent sample *t*-test.

*Statistical significance *P*<.05, two-tailed.

**Domain fell out of + 0.5 SD of population mean.

Overview of individual word analyses

Due to the sheer volume of stimuli evaluated (194 pain-related and 194 non-pain-related words), we opted to report here the overall findings of the individual pain-related words analyses as well as the results of the linear mixed model analyses. We deliberately chose not to provide a fixed recommended database as the selection of stimuli should depend on the study's research objectives (e.g., short stimulus presentation time vs. long stimulus presentation time) and target population(s) (e.g., comparison between individuals with and without chronic pain vs. research with only individuals without chronic pain) – see guidelines. The *English Linguistic Pain Stimulus Set (ELiPSS)*, which contains the complete list of pain-related and non-pain-related words, along with response characteristics (proportion and RTs of categorisation responses, and explicit pain ratings), word length and frequency of word usage in the English language (according to SUBTLEX-US corpus [6]) as well as the results of the statistical analyses comparing responses across groups is available in Multimedia Appendix 2 and on the Open Science Framework (https://osf.io/qch6y/?view_only=1b236874b2e046a5a194eb519463225b).

Speeded categorisation responses

In the 194 individual pain-related word analyses, there were 65 (33.5%) significant categorisation differences, with all words, except one (i.e., *stomachache*), being perceived as more pain-related by individuals with chronic pain compared to those without chronic pain. There were also 22 (11.3%) RT differences, with 19 (86.4%) of these words being categorised faster by individuals with chronic pain, suggesting that individuals with chronic pain have a stronger association with these words compared to those without chronic pain. The five highest categorised words for the chronic pain group were *pain* (used as a control word), *painful*, *chronic*, *aching*, and *suffering* whereas for the non-chronic pain group the words were *pain* (used as a control word), *ache*, *headache*, *hurting*, and *agony* (see Multimedia Appendix 2).

The five lowest categorised words for the chronic pain group were *lock*, *transfixing*, *booboo*, *dependent*, and *flickering* whereas for the non-chronic pain group the words were *transfixing*, *lock*, *flickering*, *interfere*, and *dependent* (see Multimedia Appendix 2).

Table 2 displays the summary statistics for RT by group and pain word classifications and Table 3 displays the results of the linear mixed models for RT for the complete dataset. First, we investigated whether there was a difference in RT between the two groups (Model 1A). Results showed no significant main effect of group (P=.37), suggesting no overall difference in RTs between the chronic and non-chronic pain group. Next, we investigated whether there were differences in RT between the two groups across the pain word classifications. Model 2B was selected as the best model for our analyses since the log likelihood ratio tests indicated that model 2B (AIC=261663.84, BIC=261726.09) fit the data better than did model 1A (AIC=261686.65, BIC=261725.56; $X^2_{(3)}$=28.81, P=<.001) and was more parsimonious than model 1B that did not fit the data significantly better (AIC=261666.11, BIC=261751.72; $X^2_{(3)}$=3.73, P=.29). Within this model, results showed no significant main effect of group (P=.46). However, there was a significant main effect of pain word classification, such that sensory pain words were classified significantly faster as pain-related than ill-health words (P=<.001) and threat words (P=<.001). There was no significant difference in mean RT between sensory and affective pain words (P=.15).

In addition, given many words in the dataset received a low proportion of categorisation responses, it was decided to create a dataset that contained a subset of the 194 pain-related words to assess whether results differed with that of the complete dataset. We opted to include words that at least 70% of individuals with and without chronic pain categorised into the pain category (speeded word categorisation paradigm). After removing the words that did not meet the criterion, 74 words remained in the data pool (see Multimedia Appendix 2 for the list of words). Table 2 displays the summary statistics for RT by group and pain word classifications

and Table 4 displays the results of the linear mixed models for RT for the selected dataset. First, we investigated whether there was a difference in RT between the two groups (Model 1A). Results showed again no significant main effect of group (*P*=.20), suggesting no differences in RTs between the chronic and non-chronic pain group. Next, we investigated whether there were differences in RT between the two groups across the pain word classifications. Model 2B was selected as the best model for our analyses since the log likelihood ratio tests indicated that model 2B (AIC=120323.78, BIC=120379.98) fit the data better than did model 1A (AIC=120366.24, BIC=120401.36; $X^2_{(3)}$=48.46, *P*=<.001) and was more parsimonious than model 1B that did not fit the data significantly better (AIC=120327.33; BIC=120404.60; $X^2_{(3)}$=2.46, *P*=.48). Within this model, results showed no significant main effect of group (*P*=.18). However, there was a significant main effect of pain word classifications, such that sensory pain words were classified significantly faster as pain-related than affective pain words (*P*=.002), ill-health words (*P*=<.001), and threat words (*P*=<.001).

Table 2

Summary Statistics on Outcome Measures by Group and Pain Word Classifications for Linear Mixed Model Analyses

Variable	Full sample			Chronic pain			Non-chronic pain		
	Observations	M	SD	Observations	M	SD	Observations	M	SD
Reaction time (complete dataset)									
Sensory	7887	841.21	435.86	5381	835.27	435.39	2506	853.95	436.68
Affective	4051	844.59	425.37	2767	836.52	412.60	1284	861.98	451.36
Ill-health	3808	859.92	462.28	2532	858.31	456.20	1276	863.11	474.27
Threat	1967	869.59	500.35	1314	865.01	505.72	653	878.81	489.62
Total	17713	849.16	446.97	11994	843.68	443.13	5719	860.64	454.75
Reaction time (selected dataset)									
Sensory	4039	779.71	351.53	2605	769.46	330.43	1434	798.32	386.35
Affective	1359	807.41	387.99	883	798.48	366.99	476	823.97	424.12
Ill-health	2290	816.90	414.78	1479	813.25	390.01	811	823.55	456.69
Threat	617	836.12	475.60	395	836.15	505.58	222	836.07	418.04
Total	8305	798.69	386.39	5362	791.23	369.37	2943	812.27	415.32
Explicit pain rating (complete dataset)[a]									
Sensory	2864	6.03	3.14	1837	6.20	3.10	1027	5.73	3.18
Affective	1383	6.32	3.06	879	6.62	2.99	504	5.80	3.11
Ill-health	1432	5.91	3.13	916	6.04	3.09	516	5.69	3.18
Threat	770	5.49	3.12	488	5.78	3.10	282	4.99	3.10
Total	6449	6.00	3.12	4120	6.20	3.08	2329	5.64	3.16

Note. Reaction times are displayed in milliseconds.

[a] Item was rated on an 11-point Likert scale, ranging from 0 (*not related)* to 10 (*highly related*).

Table 3

Linear Mixed Models for Reaction Time (Complete Dataset)

	Investigating differences between groups				Investigating differences between groups across the pain word classifications							
	Model 1A: Main effect				Model 1B: Main effects and interaction terms				Model 2B: Removing interaction terms			
	B	SE	95% CI[a]	*P*	B	SE	95% CI[a]	*P*	β	SE	95% CI[a]	*P*
Intercept	878.29	5.59	867.48; 889.86	<.001*	867.18	8.00	850.61; 882.39	<.001*	864.81	6.46	851.97; 877.12	<.001*
Pain group	-5.95	6.72	-19.66; 6.70	.37	-8.65	9.67	-27.84; 10.70	.36	-5.24	6.84	-18.19; 9.22	.46
Secondary level (covariate)	-45.58	7.24	-59.38; -30.98	<.001*	-46.21	7.48	-60.83; -32.30	<.001*	-46.12	7.31	-60.00; -31.30	<.001*
Affective words	-	-	-	-	18.10	12.50	-6.02; 44.04	.16	9.83	6.94	-3.94; 23.61	.15
Ill-health words	-	-	-	-	17.82	14.40	-10.88; 45.41	.21	28.29	7.78	12.56; 43.63	<.001*
Threat words	-	-	-	-	28.72	16.85	-2.59; 62.62	.084	45.19	10.62	23.24; 66.37	<.001*
Affective*Pain	-	-	-	-	-12.09	15.02	-43.37; 15.76	.41	-	-	-	-
Ill-health*Pain	-	-	-	-	15.65	17.51	-18.44; 51.91	.37	-	-	-	-
Threat*Pain	-	-	-	-	24.61	20.82	-17.07; 65.57	.23	-	-	-	-
AIC	261686.65				261666.11				261663.84			
BIC	261725.56				261751.72				261726.09			

Note. AIC=Akaike Information Criterion; BIC=Bayesian Information Criterion; for pain word classification, sensory was the reference category; for group, non-chronic pain group was the reference category; for education level, tertiary level education was the reference category.

[a] Bootstrapped confidence intervals (1000 replicates).

*Statistical significance *P*<.05, two-tailed.

Table 4

Linear Mixed Models for Reaction Time (Selected Dataset)

	Investigating differences between groups				Investigating differences between groups across the pain word classifications							
	Model 1A: Main effect				Model 1B: Main effects and interaction terms				Model 2B: Removing interaction terms			
	B	SE	95% CI[a]	*P*	B	SE	95% CI[a]	*P*	β	SE	95% CI[a]	*P*
Intercept	821.25	6.84	807.53; 835.44	<.001*	801.85	9.39	784.64; 821.19	<.001*	797.12	7.68	782.27; 812.76	<.001*
Pain group	-10.58	8.39	-26.67; 5.75	.20	-17.58	11.42	-40.88; 4.60	.13	-10.30	8.16	-27.25; 5.41	.18
Secondary level (covariate)	-33.39	8.81	-51.96;-15.76	<.001*	-33.69	9.07	-51.93; -15.70	.002*	-33.62	8.88	-50.54; -15.67	<.001*
Affective words	-	-	-	-	34.69	18.37	-1.16; 72.46	.066	33.23	10.01	13.38; 52.09	.002*
Ill-health words	-	-	-	-	36.86	15.97	4.36; 68.19	.021*	48.22	9.06	30.54; 66.84	<.001*
Threat words	-	-	-	-	49.51	25.57	2.40; 103.16	.058	73.98	16.82	42.83; 108.64	<.001*
Affective*Pain	-	-	-	-	-2.18	22.30	-47.30; 43.88	.92	-	-	-	-
Ill-health*Pain	-	-	-	-	17.59	19.32	-18.91; 55.63	.37	-	-	-	-
Threat*Pain	-	-	-	-	38.22	34.45	-30.72; 104.30	.28	-	-	-	-
AIC	120366.24				120327.33				120323.78			
BIC	120401.36				120404.60				120379.98			

Note. AIC=Akaike Information Criterion; BIC=Bayesian Information Criterion; for pain word classification, sensory was the reference category; for group, non-chronic pain group was the reference category; for education level, tertiary level education was the reference category.

[a] Bootstrapped confidence intervals (1000 replicates).

*Statistical significance *P*<.05, two-tailed.

Explicit pain ratings

In the 194 individual pain-related word analyses, there were 18 (9.3%) significant explicit pain-relatedness rating differences, with all these words receiving higher ratings from individuals with chronic pain, suggesting that individuals with chronic pain tend to perceive these words as more pain-related compared to those without chronic pain. The five highest rated pain words for the chronic pain group were *pain*, *painful*, *excruciating*, *migraine*, and *agonising* whereas for the non-chronic pain group the words were *pain*, *unbearable*, *agonising*, *agony*, and *painful* (see Multimedia Appendix 2). The five lowest rated pain words for the chronic pain group were *boring*, *lock*, *germs*, *gritty*, and *flickering* whereas for the non-chronic pain group the words were *lock*, *boring*, *apprehension*, *dependent*, and *interfere* (see Multimedia Appendix 2).

Table 2 displays the summary statistics for explicit pain rating by group and pain word classifications and Table 5 displays the results of the linear mixed models for explicit pain rating for the complete dataset. First, we investigated whether there was a difference in explicit pain ratings between the two groups (Model 1A). Results showed a significant main effect of group (P=<.001), suggesting that individuals with chronic pain rated the relatedness of pain words significantly higher than those without chronic pain. Next, we investigated whether there were differences in explicit pain ratings between the two groups across the pain word classifications. Model 2B was selected as the best model for our analyses since the log likelihood ratio tests indicated that model 2B (AIC=30965.35, BIC=31019.53) fit the data better than did model 1A (AIC=31004.59, BIC=31038.45; $X^2_{(3)}$=45.24, P=<.001) and was more parsimonious than model 1B that did not fit the data significantly better (AIC=30966.14, BIC=31040.63; $X^2_{(3)}$=5.22, P=.16). Within this model, results showed a significant main effect of group (P=<.001) and a significant main effect of pain word classifications. Overall, sensory pain words received significantly lower ratings than affective pain words (P=.002), but

significantly higher ratings than threat words (*P*=<.001). There was no significant difference in pain ratings between sensory pain words and ill-health words (*P*=.35).

Table 5

Linear Mixed Models for Explicit Pain Rating (Complete Dataset)

	Investigating differences between groups				Investigating differences between groups across the pain word classifications							
	Model 1A: Main effect				Model 1B: Main effects and interaction terms				Model 2B: Removing interaction terms			
	B	SE	95% CI[a]	*P*	B	SE	95% CI[a]	*P*	β	SE	95% CI[a]	*P*
Intercept	5.62	0.05	5.51; 5.73	<.001*	5.67	0.08	5.52; 5.83	<.001*	5.64	0.07	5.51; 5.78	<.001*
Pain group	0.51	0.07	0.36; 0.65	<.001*	0.46	0.10	0.26; 0.67	<.001*	0.51	0.07	0.37; 0.64	<.001*
Secondary level (covariate)	0.28	0.08	0.14; 0.43	.002*	0.28	0.08	0.12;0.43	<.001*	0.28	0.08	0.13; 0.44	<.001*
Affective words	-	-	-	-	0.10	0.14	-0.16; 0.39	.47	0.27	0.08	0.12; 0.43	.002*
Ill-health words	-	-	-	-	0.01	0.15	-0.27; 0.32	.92	-0.08	0.08	-0.23; 0.08	.35
Threat words	-	-	-	-	-0.66	0.17	-0.99; -0.32	<.001*	-0.52	0.10	-0.72; -0.31	<.001*
Affective*Pain	-	-	-	-	0.27	0.18	-0.10; 0.59	.13	-	-	-	-
Ill-health*Pain	-	-	-	-	-0.15	0.18	-0.52; 0.18	.42	-	-	-	-
Threat*Pain	-	-	-	-	0.23	0.21	-0.20; 0.63	.29	-	-	-	-
AIC	31004.59				30966.14				30965.35			
BIC	31038.45				31040.63				31019.53			

Note. AIC=Akaike Information Criterion; BIC=Bayesian Information Criterion; for pain word classification, sensory was the reference category; for group, non-chronic pain group was the reference category; for education level, tertiary level education was the reference category.

[a] Bootstrapped confidence intervals (1000 replicates).

*Statistical significance *P*<.05, two-tailed.

Discussion

Principal findings

To our knowledge this is the first research presenting the development and preliminary evaluation of a large pool of pain-related and non-pain-related words in individuals with and without chronic pain. In the first study, we identified 21 independent studies, involving a total of 194 different pain-related words and 336 matched non-pain-related words. The words *throbbing*, *shooting* and *sharp* were the most widely used pain stimuli, which is unsurprising given studies tend to take their stimuli from the MPQ [30], with 61 out of the 78 pain descriptors present in the dataset. The word "pain", however, was not frequently used in the included studies, despite being the most basic pain term that is commonly and increasingly employed worldwide [9,12]. Following the review, a pool of 194 pain-related and matched non-pain-related words was selected for evaluation. In the second study, adults with and without chronic pain completed a speeded word categorisation paradigm and rated the pain-relatedness of a subset of pain words.

First, our dataset revealed that many words used in previous AB studies were not perceived by participants as related to pain. More specifically, for the non-chronic pain group, nearly two-thirds (n=119/194; 61%) of the words had low pain categorisation responses (i.e., below 70%). For the chronic pain group, one-third (n=66/194; 34%) of the words had pain categorisation responses below 70%. Contrary to expectation, the chronic pain group was not overall faster at categorising pain words compared to the non-chronic pain group. However, the 65 (33.5%) individual categorisation differences and 22 (11.3%) RT differences found in the dataset indicated that individuals with chronic pain categorised some words as more pain-related and faster than those without chronic pain. Together, these findings suggest that although, no overall differences in associative strength of pain words with the pain construct were found between individuals with and without chronic pain, individuals with chronic pain

tend to perceive pain words as more related to pain, with a subset of pain words showing higher associative strength with the pain construct, compared to those without chronic pain. These findings should be considered when using linguistic stimuli to investigate differential processing of pain information between individuals with and without chronic pain. Indeed, differences in categorisation and/or associative strength of (part of the) linguistic stimuli may contribute to explaining the mixed findings in the literature in this area, with some studies showing that individuals with chronic pain tend to preferentially attend toward pain-related stimuli [11,40,44] and others not [7,34,41,43].

Second, partially consistent with our predictions, sensory pain words were categorised faster, and thus suggested to have a stronger associative strength with the pain construct, than ill-health words, threat words, and affective pain words, with the latter only the case for the selected dataset (N=74 words). The findings are in line with two meta-analyses investigating pain-related AB (inferred from response latency), which have found the presence of pain-related ABs only when sensory pain words were used [11,44]. Together, these findings suggest that words that reflect the sensory characteristics of pain may facilitate AB, possibly due to their stronger associations with the pain construct. Of interest, our data also show that, overall, RTs to pain words remain large, which might explain why ABs at very early stages of information processing are not common.

Third, consistent with our expectations, the chronic pain group rated the words as more pain-related compared to the non-chronic pain group, suggesting that individuals with chronic pain tend to perceive words as more pain-related compared to those without chronic pain. This result might be explained by individuals with chronic pain using pain words more frequently due to the repeated or continued experience of pain. The finding is also in line with our earlier results showing that a higher percentage of individuals with chronic pain categorise words as pain-related and that some words are more strongly associated with the pain construct. This

relationship between our implicit and explicit measurements of relatedness corroborates with brain imaging studies, which have shown that pain-related words can activate the brain networks involved in the processing of pain and that this activation is stronger in individuals with chronic pain than in healthy controls [13,35,37]. Together, these findings suggest that pain-related words are processed differently by individuals with chronic pain.

Finally, our results indicate that sensory pain words were rated as more pain-related than threat words, but as less pain-related than affective pain words, and no differently than ill-health words. These results are difficult to interpret. It is possible that some pain classifications overlapped with other classifications (e.g., broader illness) or that some classifications were less pain-related, but more pain specific (e.g., sensory). Some words may also possess a degree of ambiguity (e.g., tender-pain vs. tender-gentleness) and cannot be uniquely categorised as pain-related or non-pain-related. For these words, the context in which they are assessed may contribute to the relatedness to pain. Thus, it is important that words are validated at an individual level to facilitate personal relatedness, which in turn may better activate personal pain schemata [11]. Our dataset is an important starting point for selecting pain-related and non-pain-related linguistic stimuli.

Limitations

This study has several limitations. First, the original recruitment target estimated for the chronic pain groups and the non-chronic pain group was not achieved. Second, the sample was relatively young, predominantly female, highly educated, and studying, limiting generalizability of the results. Future research should replicate the study with a larger and more balanced sample. Third, we recruited only individuals with chronic neuropathic pain and chronic musculoskeletal and/or fibromyalgia pain, which may limit the generalizability of results to other types of chronic pain conditions. We were also unable to differentiate among these types of pain because the two chronic pain groups were heterogeneous with a large

overlap in symptoms. Future research should demonstrate validity of the chronic pain word set in each specific population (and at an individual level) [48]. Fourth, analyses on the pain-relatedness ratings had less power because of reduced sample size. Fifth, we did not evaluate the specificity of pain words, which future studies should investigate to further guide the selection of stimuli. Sixth, the online environment may have introduced measurement variability due to potential external factors (e.g., hardware and software) and the environment (e.g., interruptions and noise). Future research may want to replicate the findings in a more controlled setting (e.g., laboratory). Seventh, while the inter-coder reliability for our word classification was high, further work is needed to standardise acceptable definitions for each of the classifications and existing word lists should be classified according to these definitions. This will help reduce the variance in the use of experimental stimuli and facilitate interpretation of results. Eighth, standardised effect sizes are unavailable for the linear mixed model and generalised linear model analyses due to the use of bootstrapping and gamma distribution with log link function, respectively. Lastly, we only used AB studies for developing the pool of linguistic stimuli. We encourage researchers to enlarge and extend the current dataset (https://osf.io/qch6y/?view_only=1b236874b2e046a5a194eb519463225b).

Recommendations for selecting pain linguistic stimuli

We provide, here, some guidelines to help researchers select the most suitable stimuli. First, researchers should demonstrate that stimuli are related to the experience of pain, and describe the method used [11]. Researchers may choose to develop a new set of stimuli or to select words from an existing dataset (like the *ELiPSS*) to validate the words in their specific population(s). Second, stimuli should be tailored to the target group(s). The selection of stimuli will depend on whether the research is comparing individuals with and without chronic pain or whether, for example, it is investigating only those without chronic pain. Third, the selection will vary depending on the research objectives. For example, stimuli with a high proportion of

pain-related categorisation responses and fast RTs may be better suited to be used in RT tasks (e.g., to investigate pain-related AB). Fourth, we recommend selecting stimuli that are related to pain, irrespective of the classification they belong to, and to, where appropriate, use the stimulus *pain* or *painful* as both groups rated these words as highly pain-related. If researchers decide not to use highly related pain descriptors, they should explain their stimulus selection criteria to help other researchers to select pain descriptors for their study. Fifth, each pain-related word should be paired and matched with a non-pain-related word for length and frequency [6]. Finally, researchers should report the characteristics of their stimuli set and ensure that they are freely and openly available to the research community for transparency and reproducibility [19]. To facilitate this, we have created the *Linguistic Materials for Pain (LMaP) Repository*, which is openly available on the Open Science Framework (https://osf.io/qch6y/?view_only=1b236874b2e046a5a194eb519463225b). Researchers who wish to add their new linguistic set to this database can fill out the corresponding form (https://osf.io/qch6y/?view_only=1b236874b2e046a5a194eb519463225b), which upon verification will be added to the database.

Supplementary materials

Multimedia Appendix 1: Identification and selection of pain-related and non-pain-related stimulus material (Study 1)

Multimedia Appendix 2: Evaluation of pain-related and non-pain-related stimulus material (Study 2)

Acknowledgments

The authors wish to thank all the participants who participated in this study and the organisations (e.g., Diabetes Australia) who assisted with advertising the study to communities with chronic pain.

Authors' contributions

All authors made contributions to the conception and design of the study protocol, which was undertaken as part of a PhD for JV under the supervision of the other authors. DVR developed the validation tasks for this study. JV conducted the literature search, the study, and analysed all study data. JV drafted the manuscript, and the other authors provided critical feedback and revisions. All authors read and approved the final version of the manuscript.

Abbreviations

AB: Attentional bias

AIC: Akaike Information Criterion

BIC: Bayesian Information Criterion

ELiPSS: English Linguistic Pain Stimulus Set

GCPS: Graded Chronic Pain Scale

LMaP Repository: Linguistic Materials for Pain Repository

MPQ: McGill Pain Questionnaire

NPQ: Neuropathic Pain Questionnaire

OP: Order of position

PROMIS: Patient-Reported Outcomes Measurement Information System

RT: Reaction time

SD: Standard deviation

WPI and SS-Scale: Widespread Pain Index and Symptom Severity Scale

References

1. Amtmann D, Kim J, Chung H, Bamer AM, Askew RL, Wu S, Cook KF, Johnson KL. Comparing CESD-10, PHQ-9, and PROMIS depression instruments in individuals with multiple sclerosis. *Rehabil. Psychol.* 59:220–229, 2014

2. Bar-Haim Y, Lamy D, Pergamin L, Bakermans-Kranenburg MJ, Van Ijzendoorn MH. Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychol. Bull.* 133:1-24, 2007

3. Batterham PJ, Sunderland M, Carragher N, Calear AL. Psychometric properties of 7- and 30-day versions of the PROMIS emotional distress item banks in an Australian adult sample. *Assessment.* 26:249-259, 2017

4. Bowler JO, Bartholomew KJ, Kellar I, Mackintosh B, Hoppitt L, Bayliss AP. Attentional bias modification for acute experimental pain: A randomized controlled trial of retraining early versus later attention on pain severity, threshold and tolerance. *Eur J Pain.* 21:112-124, 2017

5. Boyer MC, Compas BE, Stanger C, Colletti RB, Konik BS, Morrow SB, Thomsen AH. Attentional biases to pain and social threat in children with recurrent abdominal pain. *J Pediatr Psychol.* 31:209-220, 2006

6. Brysbaert M, New B. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods.* 41:977-990, 2009

7. Cardoso S, Fernandes C, Barbosa F. Emotional and attentional bias in fibromyalgia: A pilot ERP study of the dot-probe task. *Neurol Ther.* 10:1079-1093, 2021

8. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Amtmann D, Bode R, Buysse D, Choi S, Cook K, DeVellis R, DeWalt D, Fries JF, Gershon R, Hahn EA, Lai J-S, Pilkonis P, Revicki D, Rose M, Weinfurt K, Hays R. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J. Clin. Epidemiol.* 63:1179-1194, 2010

9. Chaturvedi J, Mascio A, Velupillai SU, Roberts A. Development of a lexicon for pain. *Front. digit. health.* 3:1-11, 2021

10. Coleshill MJ, Sharpe L, Colagiuri B. No evidence that attentional bias towards pain-related words is associated with verbally induced nocebo hyperalgesia: A dot-probe study. *Pain Reports.* 6: e921, 2021

11. Crombez G, Van Ryckeghem DM, Eccleston C, Van Damme S. Attentional bias to pain-related information: A meta-analysis. *Pain.* 154:497-510, 2013

12. Duncan G. The meanings of 'pain' in historical, social, and political context. *The Monist.* 100:514-531, 2017

13. Eck J, Richter M, Straube T, Miltner WHR, Weiss T. Affective brain regions are activated during the processing of pain-related words in migraine patients. *Pain.* 152:1104-1113, 2011

14. Fashler SR, Katz J. More than meets the eye: Visual attention biases in individuals reporting chronic pain. *J. Pain Res.* 7:557-570, 2014

15. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods.* 39:175-191, 2007

16. Fazio RH, Williams CJ, Powell MC. Measuring associative strength: Category-item associations and their activation from memory. *Polit. Psychol.* 21:7-25, 2000

17. Ferrari R, Russell AS. A questionnaire using the modified 2010 American College of Rheumatology criteria for fibromyalgia: Specificity and sensitivity in clinical practice. *J Rheumatol.* 40:1590-1595, 2013

18. Grühn D, Sharifian N: Lists of Emotional Stimuli. In: Emotion measurement.(Meiselman, H., Ed.), Woodhead Publishing, Duxford, England, 2016, pp. 145-164.

19. Hardwicke TE, Mathur MB, MacDonald K, Nilsonne G, Banks GC, Kidwell MC, Hofelich Mohr A, Clayton E, Yoon EJ, Henry Tessler M, Lenne RL, Altman S, Long B, Frank MC. Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *R Soc Open Sci.* 5:180448, 2018

20. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Commun Methods Meas.* 1:77-89, 2007

21. HealthMeasures: Patient-Reported Outcomes Measurement Information System (PROMIS) Health Measures. Available at: http://www.healthmeasures.net/explore-measurement-systems/promis
Accessed 21 December, 2021

22. Hughes A, Hirsch C, Chalder T, Moss-Morris R. Attentional and interpretive bias towards illness-related information in chronic fatigue syndrome: A systematic review. *Br. J. Health Psychol.* 21:741-763, 2016

23. Jones EB, Sharpe L, Todd J, MacDougall H, Nicholas M, Colagiuri B. Examining attentional biases, interpretation biases, and attentional control in people with and without chronic pain. *Pain.* 162:2110-2119, 2021

24. Krause SJ, Backonja M-M. Development of a neuropathic pain questionnaire. *Clin J Pain.* 19:306-314, 2003

25. Kroenke K, Talib TL, Stump TE, Kean J, Haggstrom DA, DeChant P, Lake KR, Stout M, Monahan PO. Incorporating PROMIS symptom measures into primary care practice—a randomized clinical trial. *J. Gen. Intern. Med.* 33:1245-1252, 2018

26. Lo S, Andrews S. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Front. Psychol.* 6:1-16, 2015

27. MacLeod C, Mathews A, Tata P. Attentional bias in emotional disorders. *J. Abnorm. Psychol.* 95:15-20, 1986

28. MacLeod C, Rutherford E, Campbell L, Ebsworthy G, Holker L. Selective attention and emotional vulnerability: Assessing the causal basis of their association through the experimental manipulation of attentional bias. *J. Abnorm. Psychol.* 111:107-123, 2002

29. Meade AW, Craig SB. Identifying careless responses in survey data. *Psychol. Methods.* 17:437-455, 2012

30. Melzack R. The McGill Pain Questionnaire: Major properties and scoring methods. *Pain.* 1:277-299, 1975

31. Mittelstädt V, Ulrich R, König J, Hofbauer K, Mackenzie IG. The influence of reward in the Simon task: Differences and similarities to the Stroop and Eriksen flanker tasks. *Atten Percept Psychophys.* 2022 Oct 31; [Epub ahead of print] DOI: 10.3758/s13414-022-02563-7

32. Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: Detecting satisficing to increase statistical power. *J. Exp. Soc. Psychol.* 45:867-872, 2009

33. Peckham AD, McHugh RK, Otto MW. A meta-analysis of the magnitude of biased attention in depression. *Depress. Anxiety.* 27:1135-1142, 2010

34. Pidal-Miranda M, González-Villar AJ, Carrillo-de-la-Peña MT. Pain expressions and inhibitory control in patients with fibromyalgia: Behavioral and neural correlates. *Front. Behav. Neurosci.* 12:1-10, 2019

35. Richter M, Eck J, Straube T, Miltner WHR, Weiss T. Do words hurt? Brain activation during the processing of pain-related words. *Pain.* 148:198-205, 2010

36. Ritter A, Franz M, Miltner WHR, Weiss T. How words impact on pain. *Brain Behav.* 9:e01377, 2019

37. Ritter A, Franz M, Puta C, Dietrich C, Miltner WHR, Weiss T. Enhanced brain responses to pain-related words in chronic back pain patients and their modulation by current pain. *Healthcare.* 4:54-68, 2016

38. Schalet BD, Pilkonis PA, Yu L, Dodds N, Johnston KL, Yount S, Riley W, Cella D. Clinical validity of PROMIS(®) depression, anxiety, and anger across diverse clinical samples. *J. Clin. Epidemiol.* 73:119-127, 2016

39. Schoth DE, Liossi C. Biased interpretation of ambiguous information in patients with chronic pain: A systematic review and meta-analysis of current studies. *Health Psychol.* 35:944-956, 2016

40. Schoth DE, Nunes V, Liossi C. Attentional bias towards pain-related information in chronic pain: a meta-analysis of visual-probe investigations. *Clin. Psychol. Rev.* 32:13-25, 2012

41. Sitges C, García-Herrera M, Pericás M, Collado D, Truyols M, Montoya P. Abnormal brain processing of affective and sensory pain descriptors in chronic pain patients. *J. Affect. Disord.* 104:73-82, 2007

42. Smith BH, Penny KI, Purves AM, Munro C, Wilson B, Grimshaw J, Chambers WA, Smith WC. The Chronic Pain Grade questionnaire: Validation and reliability in postal research. *Pain.* 71:141-147, 1997

43. Snider BS, Asmundson GJG, Wiese KC. Automatic and strategic processing of threat cues in patients with chronic pain: A modified stroop evaluation. *Clin J Pain.* 16:144–154, 2000

44. Todd J, Van Ryckeghem DML, Sharpe L, Crombez G. Attentional bias to pain-related information: A meta-analysis of dot-probe studies. *Health Psychol. Rev.* 12:419–436, 2018

45. Tomczak M, Tomczak E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends Sport Sci.* 21:19-25, 2014

46. van Laarhoven AIM, van Damme S, Lavrijsen APM, van Ryckeghem DM, Crombez G, Evers AWM. Attentional processing of itch. *Psychol. Res.* 82:876–888, 2017

47. Van Ryckeghem DM, Crombez G, Eccleston C, Liefooghe B, Van Damme S. The interruptive effect of pain in a multitask environment: An experimental investigation. *J Pain.* 13:131-138, 2012

48. Van Ryckeghem DML, Crombez G: Assessment and Measurement in Health Psychology. In: Asmundson, G.J.G. (ed): Comprehensive Clinical Psychology (Second Edition). Oxford, Elsevier, 2022, pp. 85-94

49. Van Ryckeghem DML, Crombez G, Van Hulle L, Van Damme S. Attentional bias towards pain-related information diminishes the efficacy of distraction. *Pain.* 153:2345-2351, 2012

50. Van Ryckeghem DML, Noel M, Sharpe L, Pincus T, Van Damme S. Cognitive biases in pain: An integrated functional-contextual framework. *Pain.* 160:1489–1493, 2019

51. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain.* 50:133-149, 1992

52. Weiss T, Miltner WHR, Dillmann J. The influence of semantic priming on event-related potentials to painful laser-heat stimuli in migraine patients. *Neurosci. Lett.* 340:135-138, 2003

53. Wilford J, Osann K, Hsieh S, Monk B, Nelson E, Wenzel L. Validation of PROMIS emotional distress short form scales for cervical cancer. *Gynecol. Oncol.* 151:111-116, 2018

54. Wolfe F, Brähler E, Hinz A, Häuser W. Fibromyalgia prevalence, somatic symptom reporting, and the dimensionality of polysymptomatic distress: Results from a survey of the general population. *Arthritis Care Res. (Hoboken).* 65:777-785, 2013

55. Wolfe F, Clauw DJ, Fitzcharles M-A, Goldenberg DL, Häuser W, Katz RS, Mease P, Russell AS, Russell IJ, Winfield JB. Fibromyalgia criteria and severity scales for clinical and epidemiological studies: A modification of the ACR preliminary diagnostic criteria for fibromyalgia. *J Rheumatol.* 38:1113-1122, 2011