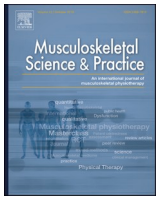
Musculoskeletal Science and Practice 73 (2024) 103150

Contents lists available at ScienceDirect

Musculoskeletal Science and Practice

journal homepage: www.elsevier.com/locate/msksp



Original article

Words matter: Effects of instructional cues on pressure pain threshold values in healthy people

David W. Evans [a], Emily Mear [b], Bradley S. Neal [b], Sally Waterworth [b], Bernard X.W. Liew [b,*]

[a] School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Edgbaston, Birmingham, United Kingdom
[b] School of Sport, Rehabilitation and Exercise Sciences, University of Essex, Colchester, Essex, United Kingdom

ARTICLE INFO

Keywords:
Pain
Quantitative sensory testing
Pain pressure threshold
Instructions
Reliability

ABSTRACT

Background: Pressure pain threshold (PPT) measurements require standardised verbal instructional cues to ensure that the increasing pressure is stopped at the correct time consistently. This study aimed to compare how PPT values and their test-retest reliability were affected by different instructional cues.
Methods: At two separate sessions, two PPT measurements were taken at the anterior knee for each of four different instructional cues: the cue of the German Neuropathic Research Network instructions ('DFNS'), the point where pressure first feels uncomfortable ('Uncomfortable'), 3/10 on the numerical pain rating scale ('3NPRS'), and where pain relates to an image from the pictorial-enhanced NPRS scale ('Pictorial'). Linear mixed modeling was used to quantify differences between pairs of instructional cues. Test-retest reliability was estimated using intraclass correlation coefficients (ICC[2,1] and ICC[2,k]).
Results: Twenty participants were recruited. The cue resulting in greatest PPT value was DFNS (394.32 kPa, 95% CI [286.32 to 543.06]), followed by Pictorial (342.49 kPa, 95%CI [248.68 to 471.68]), then Uncomfortable (311.85 kPa, 95%CI [226.43 to 429.48]), and lastly 3NPRS (289.78 kPa, 95%CI [210.41 to 399.09]). Five of six pairwise contrasts were statistically significant. Regardless of the cues, the point estimates of ICC (2,1) ranged from 0.80 to 0.86, and the ICC (2,k) values ranged from 0.89 to 0.93. No statistically significant differences were found between any pairwise contrasts of reliability indices.
Conclusion: Words matter when instructing people when to stop testing in pressure algometry. Clinicians should use the same instructional cue when assessing pain thresholds to ensure reliability.

1. Introduction

Assessing the sensitivity of body tissues in response to mechanical pressure is a fundamental part of the clinical examination of someone with pain (den Bandt et al., 2019). Pain thresholds are a commonly used measure within quantitative sensory testing (QST) paradigms (Rolke et al., 2006). The pressure pain threshold (PPT) is the minimum quantity of pressure required to induce a painful sensation when applied to a particular body site (Fischer, 1987). The most frequently employed method to measure a pressure pain threshold involves continuously increasing the magnitude of stimulus (usually at a constant rate) until pain is evoked; this is known as the ascending method of limits (Liew et al., 2021).

Prior to the commencement of PPT testing, standardised verbal instructions are provided to participants so that they can signal the assessor to stop increasing pressure (Rolke et al., 2006). It is therefore important that the instructions given to participants are clear and result in the cessation of PPT testing at the correct time. The German Neuropathic Research Network (DFNS) has produced standardised instructions for PPT (see Table 1) (Rolke et al., 2006), which have since been used in multiple studies (e.g. (Konopka et al., 2012; Liu et al., 2022; Maier et al., 2010; White et al., 2022)). Despite such attempts to standardise PPT instructions, a variety of cues are still used. These include asking participants to indicate when the applied pressure "reaches a point where it first feels uncomfortable" (O'Sullivan et al., 2014), and instructing participants to press the button when the sensation changes from "pressure to pain" (Fernández-de-las-Peñas et al., 2011) without the magnitude of pain being reported.

The definition of the presence or absence of pain is likely to be important during PPT testing and different clinical conditions determine the incidence of pain in variable ways. For example, the definition of an episode of low back pain (LBP) requires a minimum pain intensity of 2/

* Corresponding author.
E-mail addresses: bl19622@essex.ac.uk, liew_xwb@hotmail.com (B.X.W. Liew).

https://doi.org/10.1016/j.msksp.2024.103150
Received 19 March 2024; Received in revised form 25 June 2024; Accepted 21 July 2024
Available online 23 July 2024

Table 1
Instructional cues used for PPT testing.

Label	Verbatim cues
DFNS	"This procedure tests your ability to feel pressure pain above muscles. I will press this pressure measuring device against one of your muscles. Please immediately say "***NOW***" as soon as the usual sensation of pressure changes towards an additional sensation of "***burning***", "***stinging***", "***drilling***" or "***aching***"."(Rolke et al., 2006)
Uncomfortable	"Pressure will be applied at a gradual rate. Allow the pressure to increase until it reaches a point where it first feels uncomfortable and then press the button."(O'Sullivan et al., 2014)
3NPRS	"Pressure will be applied at a gradual rate. On a scale of 0–10, where 0 is no pain and 10 is the pain as bad as it could be, allow the pressure to increase until it reaches a point where you begin to feel pain with an intensity of 3/10, and then press the button."
Pictorial	"Here you will see a pain rating scale from 0 to 10. Pressure will be applied at a gradual rate. Allow the pressure to increase until it reaches a point where you begin to feel pain with an intensity matching this "  ", and then press the button. (Polomano et al., 2016)"

10 on an 11-point numerical pain rating scale (NPRS), where 0 is no pain and 10 reflects "pain as bad as it could be" (Dionne et al., 2008). In contrast, patellofemoral pain requires no minimum NPRS threshold, focusing instead on the location of pain (at, around, or behind the patella) and aggravating factors (loading a flexed knee) (Crossley et al., 2016). 'Units' of pain intensity may be calibrated differently by specific people (Walton et al., 2018), and if used to anchor PPT measurements, variation may be introduced. For instance, the label "uncomfortable" has previously been anchored to a pain score of 3/10 (Adeboye et al., 2021). Facial expression recognition scales for adults have anchored a 'sad' face to NPRS scores ranging from 3 to 4/10 (Polomano et al., 2016) to 6/10 (Garra et al., 2013).

The aim of this study was therefore to compare PPT values attained using four different instructional cues. A secondary aim of the study was to compare the test-retest reliability of PPT values derived from each instructional cue. We hypothesise that: H1) PPT values will be statistically and clinically significantly different using different instructional cues; H2) test-retest reliability of PPT will be greater using an NPRS anchored to a specific numeric threshold, compared to the DFNS cue. If instructional cues could cause a clinically meaningful alteration in PPT values, then this means that standardisation of verbal instructions is needed for PPT testing. Otherwise, a change in PPT values assumed to be caused by alterations in a participant's sensitivity may instead be driven by systematic variations in the instructional cues used during testing.

2. Methods

Prior to recruitment and data collection, ethical approval was gained from the University of University of Essex's Research Ethics Committee and the study was conducted in accordance with ethical standards laid down in the Declaration of the Helsinki. Written informed consent was gained from all participants before study enrolment. All data were collected between October 2023 and February 2024.

2.1. Participants

Healthy adults were recruited from the student and staff population at the University of Essex. The inclusion criteria were.

- No history of musculoskeletal pain requiring healthcare within the preceding 3 months
- No musculoskeletal pain at the time of testing

The exclusion criteria were.

- Inability to understand and follow instructions in verbal and written English
- Any health condition potentially causing sensory deficits, such as diabetes mellitus or a neurological disorder
- Currently taking medication that can affect sensation
- Currently pregnant
- Any history of chemotherapy
- Terminal illness with short life expectancy

Participants were asked to limit their intake of caffeine, alcohol, and any medication that could cause sleepiness or analgesia for the 24-h prior to each testing session.

2.2. Sample size

Based on a minimal detectable change (MDC) score of 158 kPa and a standard deviation of 167 kPa (Srimurugan Pratheep et al., 2018), the present study was powered to detect an effect size of 0.94. This MDC score has been similarly reported in other PPT studies (Balaguier et al., 2016; Walton et al., 2011). Based on a paired-sample *t*-test, with an alpha of 0.05, and a power of 0.8, a minimum sample size of 18 was needed. To allow for potential dropouts or data loss, 20 participants were recruited.

2.3. Data collection

Descriptive characteristics of age, height, body mass, sex, ethnicity, and time lag between test sessions were recorded. Participants attended two data collection sessions at a university laboratory, each separated by a minimum of 48 hours (Middlebrook et al., 2020).

All PPT tests were conducted by a single novice examiner (EM) with minimal experience in PPT testing prior to the study. The examiner was trained (~2 h of practice) to use the algometer by a supervising experienced researcher (BL). A hand-held digital pressure algometer (Medoc Ltd, Israel) with a contact probe sized 1 cm^2 was used for all PPT testing, which was undertaken unilaterally at the anterior knee of the dominant leg (defined as the side used to kick a ball), 2 cm proximal to the midpoint of the superior edge of the patella (Srimurugan Pratheep et al., 2018). The algometer probe was positioned perpendicularly to the skin during all PPT measurements (Rolke et al., 2006). A loading rate of 30 kPa/s, based on real-time on-screen applied force feedback, was used (Balaguier et al., 2016; Srimurugan Pratheep et al., 2018).

During PPT testing, participants sat in an upright supine position, with their legs extended. Participants were asked to press a button with the hand, on the side of the tested leg, at the moment indicated by the instructional cue being used (Table 1). Before recording PPT values, participant familiarization with the algometer was provided by twice applying pressure to their dominant forearm extensor muscles.

For each of the four instructional cues (Table 1), two repetitions of PPT testing were completed, with the order of the different cues randomised to avoid order effects (Liew et al., 2021; Middlebrook et al., 2020). If the participant failed to report pain at the level referred to within a given cue the test would be stopped at an application of 1000 kPa pressure for safety purposes, with this value recorded as the PPT (Liebano et al., 2011). To avoid tissue injury (e.g., bruising) (Graven-Nielsen et al., 2012; Nie et al., 2009; Ohrbach and Gale, 1989) and temporal sensitization via 'wind-up' (Chesterton et al., 2007; Srimurugan Pratheep et al., 2018), a 1-min interval was observed between consecutive PPT assessments (Balaguier et al., 2016).

2.4. PPT instructional cues

Four different instructional cues (Table 1) were presented in randomised order. Cue one was the widely used DFNS cue, which was verbal only. Cue two was another verbal-only cue, where pressure first felt uncomfortable, which we termed 'Uncomfortable' (O'Sullivan et al.,

2014). Cue three was also verbal only and incorporated a threshold of 3/10 from the NPRS and was labelled '3NPRS'. For cue four, an image from a pictorial enhanced NPRS scale (Polomano et al., 2016) (Fig. 1) that aligned with the 3/10 threshold of cue three was used, and this cue was labelled 'Pictorial'. This image was printed and positioned such that the participant could see it throughout all PPT testing.

2.5. Statistical analysis

All analyses were performed using R software (v4.3.0). The average of two consecutive PPT trials at the same site was taken and used for all analyses. A generalised linear mixed model (*lme4* package v1.1–35.1) was performed using a Gaussian distribution with a log-link function, with the PPT value as the outcome, instructional cues as the independent variable, and two random intercepts: for different participants and different sessions. A log-link function was used because an exploratory analysis suggested a non-normal distribution pattern of the PPT values. Pairwise contrasts (*emmeans* package v1.9.0) were then performed to quantify any differences between each pair of instructional cues. PPT values were back-transformed to their original scale for reporting. Statistically significant contrasts were defined when the 95% confidence interval (CI) does not contain the zero value.

Test-retest relative reliability between the two sessions was evaluated using the intraclass correlation coefficient (ICC[2,1] and ICC[2,k]; *SimplyAgree* package v0.2.0) based on their respective formulas from (Shrout and Fleiss, 1979). ICCs were calculated for each of the four instructional cues. For test-retest reliability in a clinical setting, an ICC (2,1) can be interpreted as "the stability of a single PPT test value across repeated sessions". Also, an ICC(2,k) can be interpreted as "the stability of an averaged PPT test value across random blocks of sessions". The relevance of an ICC(2,k) in test-retest scenarios is that if the reliability of PPT testing from a single session was poorer, one could consider taking the average PPT value of two to three sessions closely in time (such as in a research or inpatient setting). Differences in ICCs between cues and their 95% CI were calculated using bootstrapped resampling (B = 1000). The following thresholds were used to interpret ICC values: poor (<0.50), moderate (0.5–0.75), good (>0.75 to 0.90), and excellent (>0.90) reliability (Koo and Li, 2016). The test-retest absolute reliability was assessed using the standard error of measurement (SEM) ($SEM = SD\sqrt{1 - ICC_{2,1}}$) and the Bland-Altman limits of agreement (LOA) (Bland and Altman, 1986). SEM (in units of kPa) reflects the measurement error if a participant's PPT test value intrinsically varies without a systematic change in the participant across sessions. SD represents the average standard deviation of the PPTs across two sessions. The LOA was used to assess the disagreement between the PPTs across the two sessions.

3. Results

The descriptive characteristics of all participants are shown in Table 2. The mean of the within-trial loading rate was 29.1k Pa/s, whilst the associated standard deviation was 19.9 kPa/s. The safety limit of 1000 kPa was reached in four individual PPT measurements: two from the same participant with the Pictorial cue at the 2nd session; and one from each of the two participants with the DFNS cue during session one and two respectively.

On the transformed logarithmic scale, our mixed model produced an intercept (DFNS cue) of 6.01 (95%CI 5.80 to 6.21, t = 57.44, P < 0.001), and the coefficients of $\beta_{Uncomfortable}$ −0.21 (95%CI -0.29 to −0.14, t = −5.52, P < 0.001), β_{3NPRS} −0.31 (95%CI -0.39 to −0.23, t = −7.61, P < 0.001), and $\beta_{Pictorial}$ −0.14 (95%CI -0.21 to −0.07, t = −3.74, P < 0.001). On the original scale, in order of greatest to smallest magnitude of mean PPT value, the DFNS produced the greatest value (406.36 kPa, 95%CI [331.04 to 498.80]), followed by Pictorial (353.77 kPa, 95%CI [287.72 to 434.99]), then Uncomfortable (328.36 kPa, 95%CI [266.78 to 404.17]), followed lastly by 3NPRS (297.39 kPa, 95%CI [241.15 to

Table 2
Descriptive characteristics of study sample.

Characteristics	
Sex	11 female/9 male
Age (years old)[a]	26.8 (5.3)
Body mass index (kg/m^2) [a]	25.9 (3.4)
Leg dominance	18 right/2 left
Ethnicity	
• English, Welsh, Scottish, Northern Irish or Britain	14
• Any other Black, Black British or Caribbean background	2
• White and Asian	1
• White and Black African	1
• African	1
• Any other mixed or multiple ethnic background	1
Lag between two test sessions (days) [a]	2(0)

[a] Numbers reflect mean (one standard deviation).

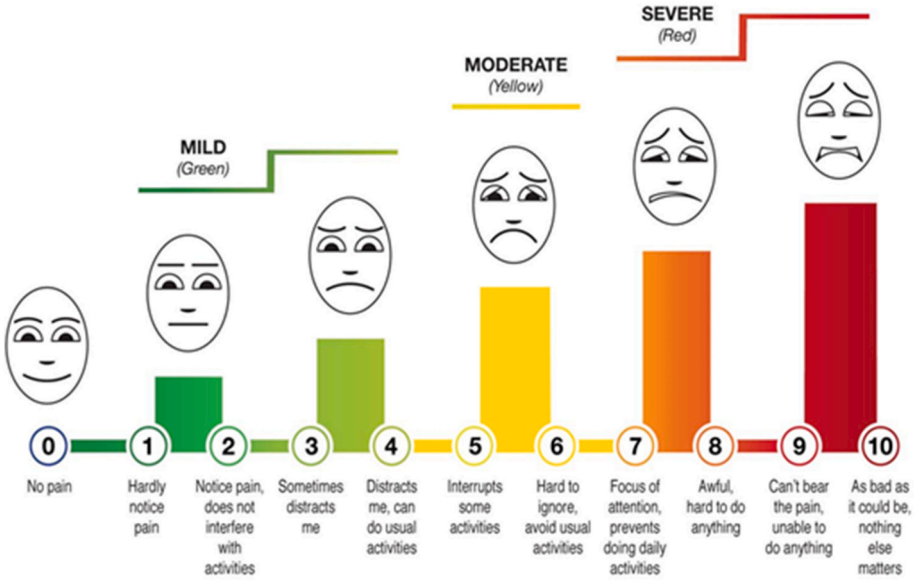


Fig. 1. Pictorial pain rating scale (Polomano et al., 2016).

366.74]) (Fig. 2).

Pairwise contrasts found five out of six comparisons to be statistically significant: the DFNS cue resulted in 78.0 kPa (95%CI 46.54 to 109.45), 109.0 kPa (95%CI 74.08 to 143.86), and 52.6 kPa (95%CI 23.25 to 81.92) greater PPT values than Uncomfortable, 3NPRS, and Pictorial, respectively (Fig. 2). The Uncomfortable cue resulted in 31.0 kPa (95% CI 3.09 to 58.86) greater PPT value than the 3NPRS cue, while the 3NPRS cue resulted in 56.4 kPa (95%CI 26.97 to 53.07) smaller PPT value than the Pictorial (Fig. 2).

Regardless of the instructional cue used, the point estimates of our ICC (2,1) values ranged from 0.80 to 0.86, and those of the ICC (2,k) values ranged from 0.89 to 0.93 (Fig. 3). ICC (2,k) values indicated that all cues resulted in uncertainty values completely above the threshold of good reliability (ICC >0.75) (Fig. 3). For ICC (2,1) the lower boundary of the 95%CI of all cues crossed below the threshold of 0.75 into the moderate reliability category (Fig. 3). No statistically significant differences were found between any pairwise contrasts of cues for either reliability index, with differences as small as 0.01 (95%CI -0.32 to 0.23) to a difference as large as 0.09 (95%CI -0.13 to 0.36) (***see Supplementary material***). The SEM ranged from 70.45 kPa (95%CI 45.45 to 87.09) to 84.87 kPa (95%CI 47.44 to 133.61), where the 3NPRS cue resulted in the smallest and the DFNS in the largest SEM error (Table 3). Fig. 4 presents the Bland-Altman plots for the between sessions difference for each of the four cues used (Table 3). Difference values outside the LOA indicate a real difference in PPT values between sessions (Fig. 4).

4. Discussion

To our knowledge, this is the first study to evaluate the effect of different instructional cues on obtained values of PPT. Our findings underscore the important role that instructional cues play in gaining accurate and reliable PPT measurements. We found statistically significant differences between absolute PPT values gained using different instructional cues, where the PPT was the greatest using the DFNS cue, and lowest with the 3NPRS cue, even when taking the order of testing into account through randomisation.

Previous studies have reported MDC values ranging from 92.9 to 196.3 kPa at the knee (Srimurugan Pratheep et al., 2018), 113.4–154.4 kPa at the neck (Walton et al., 2011), and 75–235 kPa at the low back (Balaguier et al., 2016). Hence, the differences in mean PPT values obtained with different cues may exceed the MDC values, and thus indicate systematic alterations beyond intrinsic PPT variation. The magnitude of differences between cues may be clinically important, given that a change of 83 kPa in PPT value at the neck was able to correctly

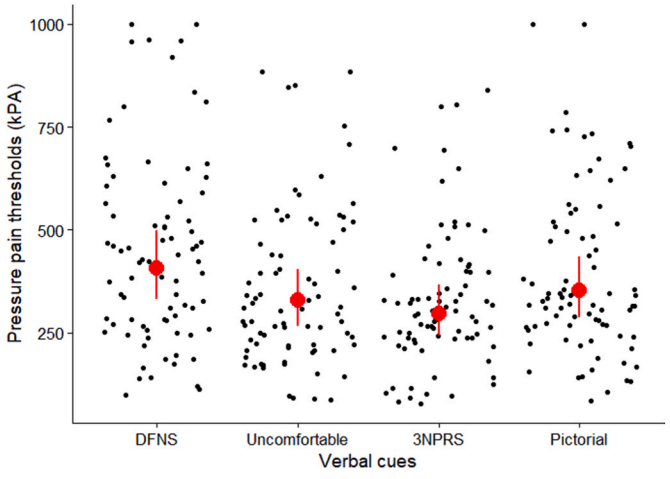


Fig. 2. Modelled marginal means with 95% CI (red) alongside individual raw data points (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

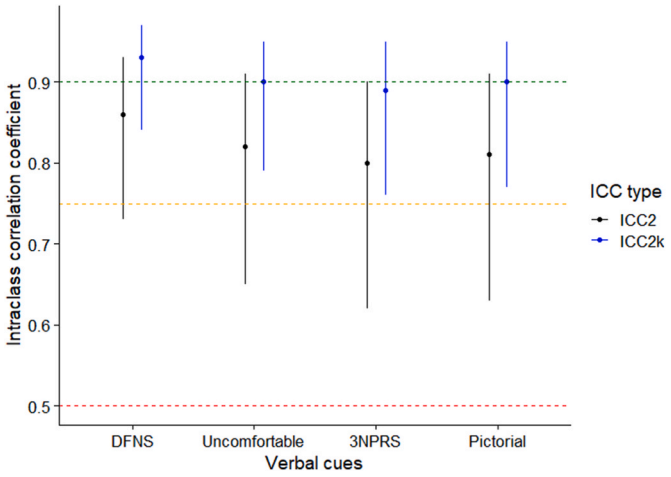


Fig. 3. Point estimate with 95%CI of Intraclass correlation coefficient values across the four verbal cue conditions. Green dashed line indicates the threshold for excellent reliability, orange dashed line for good reliability, and red dashed line for moderate reliability. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Point estimate (95% confidence interval) of the Test-retest absolute reliability indicators.

Cue conditions	SEM (kPa)	Bias, 95% LOA (kPa)
DFNS	84.87 (47.44, 133.61)	22.71 (−212.53, 257.96)
Uncomfortable	81.82 (36.17, 117.00)	−23.16 (−249.94, 203.62)
3NPRS	70.45 (45.45, 87.09)	35.72 (−230.99, 159.55)
Pictorial	84.34 (49.15, 112.21)	−44.02 (−277.79, 189.75)

Abbreviations: SEM – standard error of measurement, LOA – limits of agreement.

discriminate with moderate diagnostic accuracy, individuals who did or did not improve in their neck pain symptoms (Walton et al., 2014). The DFNS cue resulted in the greatest mean PPT value (i.e., lowest pain sensitivity). This verbal cue has been widely used in previous studies (e.g. (Konopka et al., 2012; Liu et al., 2022; Maier et al., 2010; White et al., 2022)) and asks participants to simultaneously monitor the onset of four sensations: burning, stinging, drilling, and aching (Rolke et al., 2006). The required cognitive workload of the DFNS cue could have led to a delay in response, which could, in turn, manifest as a higher PPT value. Previous research has demonstrated that attentional focus can affect the values of pain thresholds, which supports this possibility (Chayadi andMcConnell, 2019; Hoegh et al., 2019; Meyers et al., 2023).

The instructional cue that produced the lowest mean PPT value (i.e., greatest pain sensitivity) was the 3NPRS, which was explicitly anchored onto a 3/10 NPRS score. Logically, any score above zero on the NPRS is a pain response; hence, one would expect a NPRS score of 3/10 to produce a significantly higher PPT score than if anchored to a NPRS score of 1/10. It is unlikely that temporal summation affected the present PPT findings, even though we may be inducing pressure to “create” pain up to an NPRS intensity of 3/10. This was because to create temporal summation, repeated pressure would have to be applied at a much higher frequency (using a testing interval of 1 s or less), and a prior study reported that temporal summation did not happen even with a testing interval of 30 s (Nie et al., 2006). The reference score of 3/10 was specifically chosen to match the facial expression picture used within the Pictorial cue, but the PPT value from this cue was significantly larger than the value from the 3NPRS cue. At the very least, this suggests that further work needs to be performed in matching facial images representing pain responses to NPRS scores.

It is possible that asking participants to consider a 3/10 pain

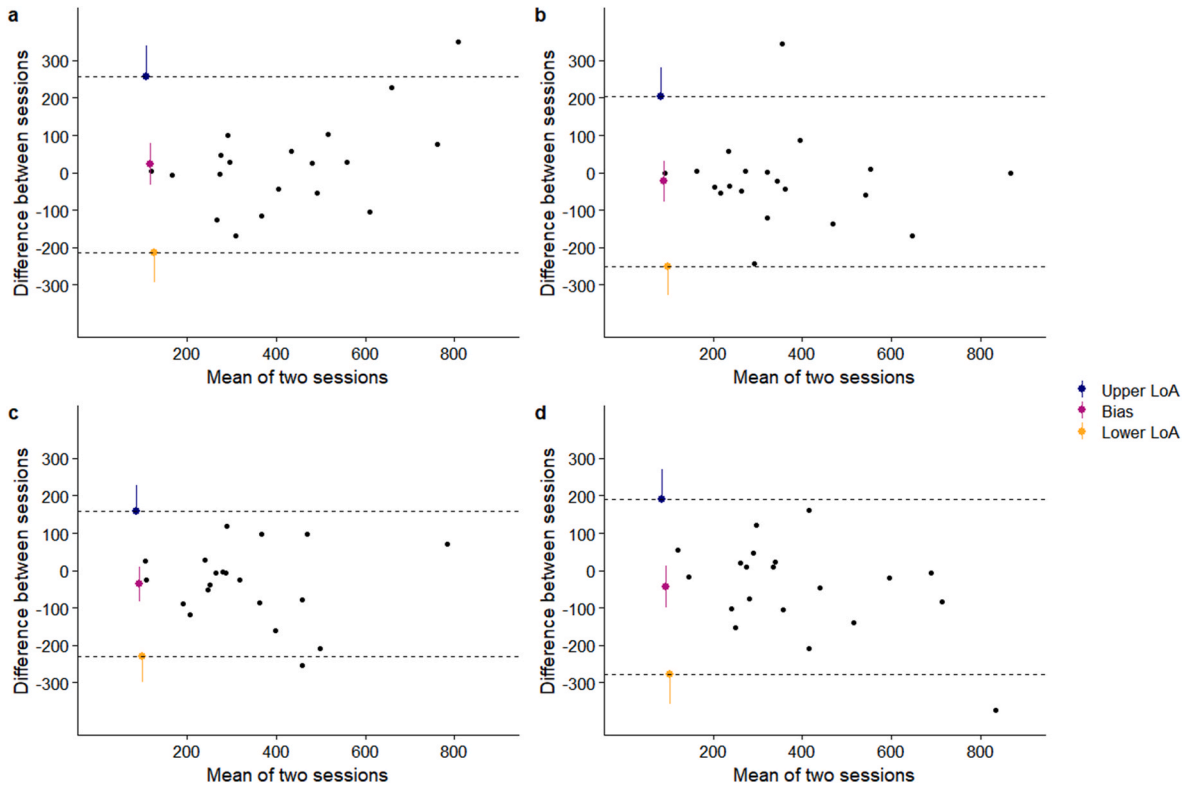


Fig. 4. Bland-Altman plots for the following cues: a) DFNS, b) Uncomfortable, c) 3NPRS, and d) Pictorial. Dashed lines (—) indicate the upper and lower 95% limits of agreement. Purple point represents the point estimate of the mean bias (95% confidence interval). Blue point represents the point estimate of the upper limit of agreement (with error bar reflecting upper boundary of 95% confidence interval). Orange point represents the point estimate of the lower limit of agreement (with error bar reflecting lower boundary of 95% confidence interval). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

experience primes them to expect a more painful experience and, in doing so, causes them to react sooner to the pressure stimulus. Cognitive processing biases are known to exist in people with chronic pain (Pincus and Morley, 2001) and so were not to be expected in the group of healthy people forming this study sample. Nevertheless, previous experiments in healthy people (Moseley and Arntz, 2007) have shown that innocuous contextual factors have the potential to bias individuals toward reporting more painful responses. It is possible that a NPRS score of 3/10 may be perceived by participants as a lower intensity than anticipated, when compared to the notions of pain being 'uncomfortable' or causing pain that is "burning", "stinging", "drilling" or "aching". Speculatively, the intensity of pain required to produce an emotional facial expression might be greater than the first moment of pain onset. Further investigations are required to unpack these findings.

Only the DFNS cue's point estimate of ICC[2,k] crossed a threshold of >0.90, indicating excellent reliability. However, the 95%CI uncertainty values of all four cues' ICC[2,k] suggest that PPT testing at the knee had good to excellent test-retest reliability. The uncertainty values of our ICC [2,1] values indicate that if reliability was taken from a single session, PPT testing at the knee had moderate to excellent reliability. Our comparison of the point estimates of the mean PPT between instructional cues cannot provide evidence for the support of one cue over another but does challenge assumptions about the construct validity of PPTs. It would seem sensible to align the cues of PPT testing to the definition of specific musculoskeletal pain disorders. Notably, systematic reviews of PPT studies do not currently extract information about which instructional cues were used (Amiri et al., 2021). Given the differences obtained by different cues that we have found in both absolute PPT values and their reliability, we would strongly recommend that future studies and systematic reviews include details of this important factor.

The present study has clinical and scientific implications and calls for greater standardisation of the instructions used in PPT testing. Given previously reported MDC values ranging from 75 to 235 kPa (Balaguier et al., 2016; Srimurugan Pratheep et al., 2018; Walton et al., 2011), variation in the verbal instructional cues could cause a variation in PPT values beyond the intrinsic noise of PPT testing. The magnitude of difference in PPT values caused by differences in verbal cues may be clinically meaningful (Walton et al., 2014). This means that alterations in PPT values overtime thought to be caused by an alteration in the health of the participant may be caused by variations in verbal cues used in PPT testing.

This study has several limitations. First, we powered this study based only on a single effect size, and not on the entire parameter space of our included mixed model. Second, PPT was tested only at a single site (anterior knee). Future studies should test PPT instructional cues at multiple sites. Third, we used a single rater and so could not assess inter-rater reliability. Previous studies have demonstrated that, although inter-rater reliability of PPT is very high, some variation does exist (Bhattacharyya et al., 2023; Chesterton et al., 2007; Waller et al., 2015). It is possible that some of this inter-rater variance could be explained by how different raters verbalise instructions, which warrants future examination. Fourth, our data are also limited by being drawn from asymptomatic people, which limits the generalisability of our findings to clinical populations. In particular, the emotional, cognitive, and sensory state of an individual in pain is likely to affect their pain sensitivity and corresponding PPT values (Carriere et al., 2019; Lalouni et al., 2021; Meints et al., 2019; Steinmetz et al., 2023), potentially confounding PPT values using a pictorial scale to the greatest extent. Future research should investigate the application of these findings in diverse patient populations, including those with pain conditions, to validate the broader applicability of standardized instructional cues in enhancing the accuracy and reliability of PPT measurements.

5. Conclusions

Words matter in the type of instructional cues used in PPT testing. Different instructional cues produce significantly different PPT values, although substantial inter-session reliability was not affected. Using identical instructional cues when re-testing facilitates reliability, irrespective of the instructional cue used.

Funding source

None.

CRediT authorship contribution statement

David W. Evans: Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Conceptualization. **Emily Mear:** Writing – review & editing, Writing – original draft, Project administration, Investigation, Formal analysis, Data curation. **Bradley S. Neal:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Data curation, Conceptualization. **Sally Waterworth:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Conceptualization. **Bernard X.W. Liew:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors have no conflicts of interest to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.msksp.2024.103150.

References

Adeboye, A., Hart, R., Senapathi, S.H., Ali, N., Holman, L., Thomas, H.W., 2021. Assessment of functional pain score by comparing to traditional pain scores. Cureus 13, e16847.
Amiri, M., Alavinia, M., Singh, M., Kumbhare, D., 2021. Pressure pain threshold in patients with chronic pain: a systematic review and meta-analysis. Am. J. Phys. Med. Rehabil. 100.
Balaguier, R., Madeleine, P., Vuillerme, N., 2016. Is one trial sufficient to obtain excellent pressure pain threshold reliability in the low back of asymptomatic individuals? A test-retest study. PLoS One 11, e0160866.
Bhattacharyya, A., Hopkinson, L.D., Nolet, P.S., Srbely, J., 2023. The reliability of pressure pain threshold in individuals with low back or neck pain: a systematic review. Br. J. Pain 17, 579–591.
Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1, 307–310.
Carriere, J.S., Martel, M.O., Meints, S.M., Cornelius, M.C., Edwards, R.R., 2019. What do you expect? Catastrophizing mediates associations between expectancies and pain-facilitatory processes. Eur. J. Pain 23, 800–811.
Chayadi, E., McConnell, B.L., 2019. Gaining insights on the influence of attention, anxiety, and anticipation on pain perception. J. Pain Res. 12, 851–864.
Chesterton, L.S., Sim, J., Wright, C.C., Foster, N.E., 2007. Interrater reliability of algometry in measuring pressure pain thresholds in healthy humans, using multiple raters. Clin. J. Pain 23, 760–766.
Crossley, K.M., Stefanik, J.J., Selfe, J., Collins, N.J., Davis, I.S., Powers, C.M., Callaghan, M.J., 2016. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 1: terminology, definitions, clinical examination, natural history, patellofemoral osteoarthritis and patient-reported outcome measures. Br. J. Sports Med. 50, 839–843.
den Bandt, H.L., Paulis, W.D., Beckwée, D., Ickmans, K., Nijs, J., Voogt, L., 2019. Pain mechanisms in low back pain: a systematic review with meta-analysis of mechanical quantitative sensory testing outcomes in people with nonspecific low back pain. J. Orthop. Sports Phys. Ther. 49, 698–715.
Dionne, C.E., Dunn, K.M., Croft, P.R., Nachemson, A.L., Buchbinder, R., Walker, B.F., Von Korff, M., 2008. A consensus approach toward the standardization of back pain definitions for use in prevalence studies. Spine (Phila Pa 1976) 33, 95–103.
Fernández-de-las-Peñas, C., Ortega-Santiago, R., Cuadrado, M.L., López-de-Silanes, C., Pareja, J.A., 2011. Bilateral widespread mechanical pain hypersensitivity as sign of central sensitization in patients with cluster headache. Headache J. Head Face Pain 51, 384–391.
Fischer, A.A., 1987. Pressure algometry over normal muscles. Standard values, validity and reproducibility of pressure threshold. Pain 30, 115–126.
Garra, G., Singer, A.J., Domingo, A., Thode Jr., H.C., 2013. The Wong-Baker pain FACES scale measures pain, not fear. Pediatr. Emerg. Care 29, 17–20.
Graven-Nielsen, T., Wodehouse, T., Langford, R.M., Arendt-Nielsen, L., Kidd, B.L., 2012. Normalization of widespread hyperesthesia and facilitated spatial summation of deep-tissue pain in knee osteoarthritis patients after knee replacement. Arthritis Rheum. 64, 2907–2916.
Hoegh, M., Seminowicz, D.A., Graven-Nielsen, T., 2019. Delayed effects of attention on pain sensitivity and conditioned pain modulation. Eur. J. Pain 23, 1850–1862.
Konopka, K.-H., Harbers, M., Houghton, A., Kortekaas, R., van Vliet, A., Timmerman, W., van Wijhe, M., 2012. Bilateral sensory abnormalities in patients with unilateral neuropathic pain; A quantitative sensory testing (QST) study. PLoS One 7, e37524.
Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15, 155–163.
Lalouni, M., Fust, J., Vadenmark-Lundqvist, V., Ehrsson, H.H., Kilteni, K., Birgitta Jensen, K., 2021. Predicting pain: differential pain thresholds during self-induced, externally induced, and imagined self-induced pressure pain. Pain 162, 1539–1544.
Liebano, R.E., Rakel, B., Vance, C.G.T., Walsh, D.M., Sluka, K.A., 2011. An investigation of the development of analgesic tolerance to TENS in humans. Pain 152, 335–342.
Liew, B., Lee, H.Y., Rügamer, D., De Nunzio, A.M., Heneghan, N.R., Falla, D., Evans, D. W., 2021. A novel metric of reliability in pressure pain threshold measurement. Sci. Rep. 11, 6944.
Liu, K., Cui, X., Zhi, M., Zhang, M., Zhao, T., Gao, X., Zhu, B., 2022. Caffeine attenuates electroacupuncture effect on pressure pain threshold and tolerance in healthy individuals: a randomized controlled trial. Front. Neurol. 13, 859624.
Maier, C., Baron, R., Tölle, T.R., Binder, A., Birbaumer, N., Birklein, F., Treede, D.R., 2010. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): somatosensory abnormalities in 1236 patients with different neuropathic pain syndromes. Pain 150, 439–450.
Meints, S.M., Mawla, I., Napadow, V., Kong, J., Gerber, J., Chan, S.T., Edwards, R.R., 2019. The relationship between catastrophizing and altered pain sensitivity in patients with chronic low-back pain. Pain 160, 833–843.
Meyers, E., Alves, M., Teugels, A., Torta, D.M., 2023. No evidence that working memory modulates the plasticity of the nociceptive system, as measured by secondary mechanical hypersensitivity. J. Pain 24, 1931–1945.
Middlebrook, N., Heneghan, N.R., Evans, D.W., Rushton, A., Falla, D., 2020. Reliability of temporal summation, thermal and pressure pain thresholds in a healthy cohort and musculoskeletal trauma population. PLoS One 15, e0233521.
Moseley, G.L., Arntz, A., 2007. The context of a noxious stimulus affects the pain it evokes. Pain 133, 64–71.
Nie, H., Arendt-Nielsen, L., Madeleine, P., Graven-Nielsen, T., 2006. Enhanced temporal summation of pressure pain in the trapezius muscle after delayed onset muscle soreness. Exp. Brain Res. 170, 182–190.
Nie, H., Graven-Nielsen, T., Arendt-Nielsen, L., 2009. Spatial and temporal summation of pain evoked by mechanical pressure stimulation. Eur. J. Pain 13, 592–599.
O'Sullivan, P., Waller, R., Wright, A., Gardner, J., Johnston, R., Payne, C., Smith, A., 2014. Sensory characteristics of chronic non-specific low back pain: a subgroup investigation. Man. Ther. 19, 311–318.
Ohrbach, R., Gale, E.N., 1989. Pressure pain thresholds in normal muscles: reliability, measurement effects, and topographic differences. Pain 37, 257–263.
Pincus, T., Morley, S., 2001. Cognitive-processing bias in chronic pain: a review and integration. Psychol. Bull. 127, 599–617.
Polomano, R.C., Galloway, K.T., Kent, M.L., Brandon-Edwards, H., Kwon, K.N., Morales, C., Buckenmaier III, C.T., 2016. Psychometric testing of the defense and veterans pain rating scale (DVPRS): a new pain scale for military population. Pain Med. 17, 1505–1519.
Rolke, R., Baron, R., Maier, C., Tölle, T.R., Treede, D.R., Beyer, A., Wasserka, B., 2006. Quantitative sensory testing in the German research Network on neuropathic pain (DFNS): standardized protocol and reference values. Pain 123.
Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 420–428.
Srimurugan Pratheep, N., Madeleine, P., Arendt-Nielsen, L., 2018. Relative and absolute test-retest reliabilities of pressure pain threshold in patients with knee osteoarthritis. Scand J Pain 18, 229–236.
Steinmetz, A., Hacke, F., Delank, K.S., 2023. Pressure pain thresholds and central sensitization in relation to psychosocial predictors of chronicity in low back pain. Diagnostics 13.
Waller, R., Straker, L., O'Sullivan, P., Sterling, M., Smith, A., 2015. Reliability of pressure pain threshold testing in healthy pain free young adults. Scand J Pain 9, 38–41.
Walton, D.M., Elliott, J.M., Salim, S., Al-Nasri, I., 2018. A reconceptualization of the pain numeric rating scale: anchors and clinically important differences. J. Hand Ther. 31, 179–183.
Walton, D.M., Levesque, L., Payne, M., Schick, J., 2014. Clinical pressure pain threshold testing in neck pain: comparing protocols, responsiveness, and association with psychological variables. Phys. Ther. 94, 827–837.
Walton, D.M., Macdermid, J.C., Nielson, W., Teasell, R.W., Chiasson, M., Brown, L., 2011. Reliability, standard error, and minimum detectable change of clinical pressure pain threshold testing in people with and without acute neck pain. J. Orthop. Sports Phys. Ther. 41, 644–650.
White, L., Smith, A.D., Farrell, S.F., 2022. Associations between resting heart rate, resting blood pressure, psychological variables and pain processing in chronic whiplash-associated disorder: a cross-sectional study. Pain Med. 23, 1882–1890.