



Case Report



Pain scales as placebos: Can pain scales change reported pain across measurements? [☆]

Joshua M. Ackerman^{*}, Jenna Goesling, Aradhna Krishna

University of Michigan, Ann Arbor, MI, USA

ARTICLE INFO

Keywords:

Measurement effect
Subjective experience
Pain
Scales
Placebo

ABSTRACT

Various aspects of measurement scales, such as whether the scale is unipolar or bipolar, or the direction of response alternatives, can influence how people evaluate their own subjective experience. Here, we demonstrate scale effects tied to repeated measurement by examining self-reported pain. In many contexts, assessment of subjective experiences is done repeatedly, as when pain patients report their pain levels using a variety of scales. We propose that this repeated measurement can impact how people report their pain. In two studies, we find that the choice of measurement scale initially used to assess pain results in different levels of self-reported pain levels at a later assessment. These repeated scale measurement effects appear to be due, in part, to the initial scales differentially affecting participant expectations for the amount of additional pain they can bear. This work extends literature on scale effects to repeated measurement. Given that many subjective experiences besides pain are also measured repeatedly (e.g., fatigue and anxiety), our results also may have wider application to other domains of experience.

Measurement scales are used in many contexts, from psychology studies to market research to medical practice. Researchers focusing on scale design—e.g., the direction of response alternatives on scales (positive to negative or negative to positive; Rammstedt & Krebs, 2007; Tourangeau, Couper, & Conrad, 2004, 2013) or whether scales are unipolar versus bipolar (Mazaheri & Theuns, 2009)—have shown that various elements of scales themselves can impact responses. We add to this literature by showing scale effects linked to repeated measurement.

Repeated measurement contexts are not uncommon. Many human ailments, from bodily injury to cancer, create consequences such as physical damage and pain that require repeated assessment to track symptoms and gauge change over time. Although many indicators of objective bodily states (e.g., torn muscles) exist, researchers and practitioners often focus on self-reports of perceived symptom severity (e.g., physical pain) based on subjective experiences. Here, we ask if the choice of measurement scale at one assessment influences reported pain at a subsequent assessment. If so, through which psychological processes might such changes occur?

1. What influences pain measurement?

Subjective evaluations of psychological states face challenges in how precisely and reliably they represent various experiences. One such challenge is the fact that subjective states need not correspond with particular objective physical conditions. This is true with physical pain, which is inherently subjective (Auvray et al., 2010). For instance, consider studies that explore the association between measures of pain severity, such as pain scales, and measures of damage. Imaging studies in individuals reporting absence of low back pain commonly show physical conditions thought to contribute to back pain, such as disc degeneration (e.g., Brinjikji et al., 2015). Conversely, when people do report experiencing back pain, imaging data often does not show evidence of physical states such as disc degeneration (Srinivas, Deyo, & Berger, 2012). Thus, there is a clear disconnect between self-reported pain scores and physical findings, suggesting that pain is influenced by more than objective damage.

Another key challenge involves the limitations of widely used self-report tools to measure pain. These tools generally ask people to report their pain level using numeric or visual scales bounded by labels such as

[☆] This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank Yuching Lin for her assistance in collecting the data.

^{*} Corresponding author at: 350 Church Street, Ann Arbor, MI 48109-1043, USA.

E-mail address: joshack@umich.edu (J.M. Ackerman).

“no pain” to “worst pain imaginable” (Cleeland & Ryan, 1994; Fillingim, Loeser, Baron, & Edwards, 2016; Younger, McCue, & Mackey, 2009), two common scales being the visual analog scale (VAS) and the numerical rating scale (NRS). Appraising the worst possible pain one can experience requires comparison to appropriate reference points, but these are known to change depending on individual differences and contextual factors (Coghill, 2010; Cruz-Almeida, Riley III, & Fillingim, 2013; Fillingim, 2005). As with other types of evaluations, people may construct their perceptions in the moment (Bettman, Luce, & Payne, 1998), making them highly susceptible to salient features of those moments.

One such momentary salient feature is the scale itself. Studies in the measurement literature show that people respond in disparate fashion to different types of scales. For instance, researchers have examined effects of number of response alternatives on Likert scales (e.g., 5 vs. 7 vs. 9 point scales; Garner, 1960; Preston & Colman, 2000), direction of response alternatives (Rammstedt & Krebs, 2007; Tourangeau et al., 2004, 2013), unipolar versus bipolar framing (Mazaheri & Theuns, 2009), and so on. For instance, in looking at the direction of response alternatives, Krebs and Hoffmeyer-Zlotnik (2010) note that when scales go from “very important” to “not very important,” responses are relatively more positive versus when the scale labels are flipped. Hartley and Betts (2010) similarly note that when a scale begins with “clear” and ends with “unclear,” respondents report higher overall numeric ratings than when the scale is reversed. Such scale effects can stem from the use of heuristics and other low cognitive effort psychological processes (Schwarz, 1999). Less research has focused specifically on the consequences of pain scales formats. Some work has examined correlations between ratings derived from various types of pain scales, such as the VAS or the NRS, as well as horizontally or vertically oriented VAS measures (e.g., Breivik & Sklogglund, 1998; Downie et al., 1978). Results from these studies have been mixed and largely depend on the exact scales being compared and the type of pain assessed (e.g., dental pain versus non-dental pain). Consistent with the latter point, variance in responses may stem from the multidimensional nature of pain, which involves both sensory-discriminative and affective-motivational aspects (Auvray et al., 2010), with common rating scales often not distinguishing these aspects (Fillingim et al., 2016; Williams, Davies, & Chadury, 2000).

Beyond such issues, one underexplored feature of pain measurement involves its psychological consequences. In particular, we suggest that the scales used to self-report pain may have placebo-like effects on people's judgments of that pain. Although research on the psychological consequences of pain scale use is relatively scant, research on pain and placebo/nocebo effects more generally is vast (e.g., Colloca, Sigauco, & Benedetti, 2008; Tracey, 2010; Wager et al., 2004). In fact, the work on pain analgesia is one of the most established experimental examinations of placebo response (Hoffman, Harrington, & Fields, 2005). For example, placebo analgesia interventions reduce pain-related processing in the spinal cord and brain (e.g., Tracey, 2010; Wager et al., 2004), whereas interventions that promote pain anticipation heighten the frequency and strength of pain responses (e.g., Kaptchuk et al., 2006; Wager et al., 2004). Such placebo (and nocebo) effects are thought to occur because of altered expectations created through learning or mere suggestion (Colloca et al., 2008; Wager et al., 2004). Importantly, it is understood that “while placebos may provide relief, they rarely cure ... they primarily address subjective and self-appraised symptoms” (Kaptchuk & Miller, 2015, p. 8).

Building from this work, we suggest that the particular scale values used to evaluate painful experiences might produce placebo effects by influencing expectations about those experiences. That is, the scale format that individuals use to judge pain at one time point may influence judgments of subsequent pain-related events.

2. Current research

We consider the consequences of repeated measurement in the context of physical pain. In settings where pain assessment is relevant, such as doctors' offices and physical therapy visits, pain is repeatedly measured to evaluate change over time. If the use of particular scales shapes the pain experience, two effects are possible. First, differences in responses to the initial assessment of pain may appear (time 1 effects). Second, use of specific scales at one time point may carry over to influence evaluations at a later time point (time 2 effects). Regarding the latter possibility, the initial measurement of pain may act as an intervention, shaping downstream responses to a subsequent painful experience, much like how the effects of placebo substances are assessed through later measurement. To test these possibilities, in two experiments, participants first evaluated a painful experience (hand immersion in cold water) by rating their pain using one variant of a pain scale. They then repeated the experience, and all participants rated their pain using a single (novel) pain scale, allowing for simple comparisons to be made across participant groups on this second trial. These self-reported evaluations represent subjective pain *intensity* (comprising both sensory-discriminative and affective-motivational dimensions of pain). In addition, we measured pain *tolerance* (amount of time in the water) during each of these experiences.

We propose that the measurement of pain itself influences certain psychological processes associated with the pain experience, with the nature of this influence depending on the scale format. Further, we propose the implications of this influence will spill over onto subsequent measurements of pain. In Experiment 2, we examine three possible psychological mechanisms that could produce such effects, including (1) dislike of particular scales transferring to self-evaluations of pain, (2) comparison of the current experience with evaluations of past pain experiences, and (3) the format of scales shaping expectations about how much pain one can bear in subsequent experiences.

3. Experiment 1

In both of our experiments, we aimed for relatively large sample sizes given norms of the pain literature that use our type of measurement (a cold pressor task) (e.g., Foxen-Craft & Dahlquist, 2017; Leong, Cano, Wurm, Lumley, & Corley, 2015; Leventhal, Brown, Shacham, & Engquist, 1979). Sensitivity analyses are provided in the Data Analysis Plan section of each experiment. All measures, manipulations, and exclusions are reported, and data were collected before analysis. Materials, data, and code are available at: https://osf.io/f9yvr/?view_only=d2aad4662d8349de81bb78aac060c720.

3.1. Method

3.1.1. Participants

We collected data from 205 undergraduate students (112 female, $M_{\text{age}} = 18.9$), who received course credit for participating in a 3-cell between-subjects experiment.

3.1.2. Apparatus and materials

A cold pressor task was used to elicit pain. The cold pressor is a widely used paradigm in pain testing given its simplicity, reliability, and validity (Koenig et al., 2014; Modir & Wallace, 2010). The device consisted of a large bin containing water and ice, a smaller bin submerged in the larger bin with small holes allowing water to enter but preventing direct contact with ice (into which participants placed a hand), and a filter which continuously circulated water in order to prevent warming (see Appendix). Water temperature was checked with a thermometer to be within 1 degree of 33° F (average $SD = 0.34$) at the outset of each trial. Participants placed their open, non-dominant hand into the small bin up to the wrist. When they could no longer tolerate the pain, they removed their hand from the water (time from

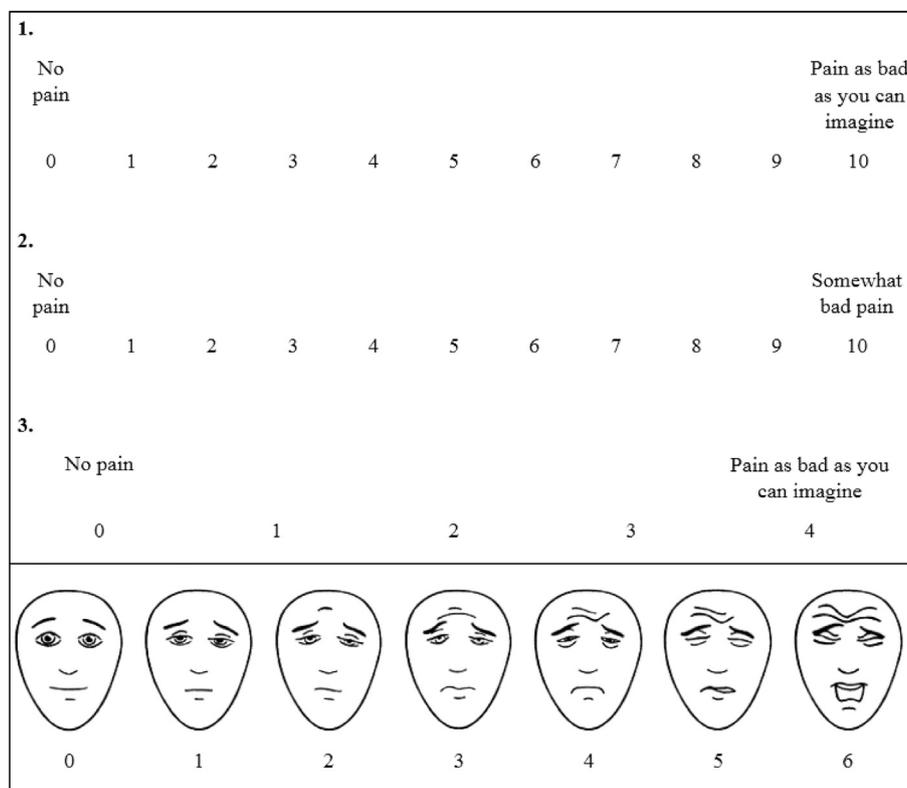


Fig. 1. Scales used in measurement of cold pressor pain in Experiments 1 and 2. Top panel: Trial 1 scales – participants received one of these depending on their condition (scale #2 was used in Experiment 1 but not in Experiment 2). Bottom panel: Trial 2 scale (all participants received this).

beginning to removal was recorded, reflecting pain tolerance). Immediately upon removal, participants were given a scale to report their level of pain (reflecting pain intensity). Two trials of this procedure were completed.

To measure pain intensity, at Trial 1, one of three scales was given depending on a randomly-assigned condition: (1) a traditional 0–10 scale (“no pain” – “pain as bad as you can imagine”), (2) a 0–10 scale with different value labels (“no pain” – “somewhat bad pain”), or (3) a 0–4 scale with traditional values labels (labeled “no pain” – “pain as bad as you can imagine”). Scale 1 is commonly used in assessment of pain (Cleeland & Ryan, 1994; Fillingim et al., 2016). Scales 2 and 3 were used to examine effects of label variation and numeric variation, respectively.

At Trial 2, a single scale was used by all participants in order to allow for direct comparison of ratings across conditions and thus afford clearer interpretation of effects: a 0–6 scale featuring affective face labels (another commonly used pain measure; Fillingim et al., 2016; Hicks, von Baeyer, Spafford, van Korlaar, & Goodenough, 2001). See Fig. 1 for depictions of pain measures and Fig. 2 for an overview of the study procedure.

3.1.3. Procedure

Following consent, participants were told that the study involved temperature sensitivity, and baseline hand temperature was measured by placing a thermometer between the thumb and middle finger of the non-dominant hand. The cold pressor task was then explained in detail. Participants were told:

Once you've put your hand in, we'd like you to leave it in for as long as you can even if it's uncomfortable. But you can take your hand out when it hurts too much to leave it in. When you take your hand out of the water, I'll immediately ask you to rate your pain verbally using a rating scale that I'll hold up for you to see.

After clarifying any questions, the participant began Trial 1 of the task

while the experimenter remained unreactive and behind the participant so as to not influence reactions. Immersion time was recorded with a stopwatch up to 4 min, at which point the task was ended if the participant's hand remained in the water (no maximum immersion time was told to participants in advance). Next, participants were shown a printed version of their assigned scale and asked to verbally report their level of pain at removal.

After a rest period of 60 s, hand temperature was measured again. To help return to baseline levels, participants then submerged their hand in a warm water bath (98°–100 °F) for the number of seconds equaling the deviation of present temperature from the individual participant's baseline temperature. Hand temperature was re-assessed 60 s after removal from the warm bath. This procedure was repeated until hand temperature returned to within 1 °F of baseline.

Trial 2 of the cold pressor task was then administered. This trial replicated the earlier one except that, after the cold pressor task, all participants received the Trial 2 pain scale (0–6 values). Following this, participants completed questionnaires assessing experience with chronic pain, current use of pain medication, mental health (General Health Questionnaire 12: Goldberg, Weisenberg, Drobkin, Blittner, & Gotestam, 1997), several psychological measures included for exploratory purposes (Perceived Vulnerability to Disease: Duncan, Schaller, & Park, 2009; Self-Monitoring: Snyder & Gangestad, 1986; Private Self-Consciousness; Scheier & Carver, 1985), and demographics. Analyses of exploratory measures are not reported here. They were then debriefed and released.

3.2. Results

3.2.1. Data analysis plan

Fifteen participants were excluded from all analyses for: failing to complete measures/procedural problems (6), cold pressor task problems (2), and hand temperatures that did not return to within ± 5 °F of baseline before Trial 2 of the cold pressor (7). Additionally, 15

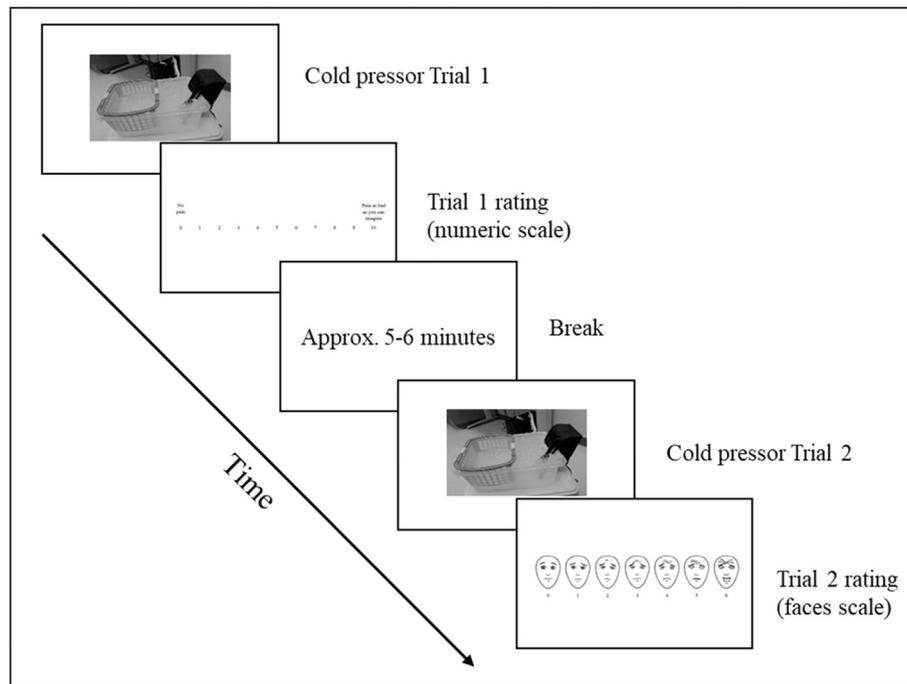


Fig. 2. Overview of the experiment procedure.

participants reported experiencing chronic pain (daily pain lasting more than three months), and 6 reported current medication use that might affect pain sensitivity. These participants were also excluded. We controlled for participant sex and age, following protocols from prior pain testing studies (e.g., Naugle & Riley 3rd, 2014; Neville et al., 2018; Racine et al., 2012). In both studies, analyses without any covariates did not change the substantive findings reported next.

For the primary analyses, we tested the effects of pain scale condition (the type of pain scale used in Trial 1) on pain intensity (self-reported pain at removal from water) and pain tolerance (total time in water) at Trial 2, controlling for sex and age. We report 95% confidence intervals for contrast differences. After exclusions, our Experiment 1 sample was sufficient to detect small-to-medium effects of $f \geq 0.241$ with 80% power in a between-subjects ANCOVA.

3.2.2. Primary analyses

For pain intensity, an ANCOVA on Trial 2 pain ratings revealed a main effect of Condition, $F(2, 164) = 3.84, p = .023, \eta_p = 0.05$, suggesting that scales used at Trial 1 impacted pain intensity at Trial 2, i.e., there was an effect of repeated pain measurement. Planned contrasts indicated that the 0–4 scale led to higher pain ratings than the traditional 0–10 scale, $p = .007, d = 0.49$ (95% CI: $-1.204, -0.195$) and marginally higher ratings than the modified 0–10 scale, $p = .081, d = 0.31$ (95% CI: $-0.922, 0.054$). No significant differences emerged between the two 0–10 scale conditions, $p = .288$, (95% CI: $-0.757, 0.227$). See Fig. 3 (Panel A) for pain values.

For pain tolerance, time measurements on Trials 1 and 2 showed acceptable skewness ($< |1.1|$ and $< |0.11|$, respectively) and kurtosis ($< |0.58|$ and $< |1.26|$, respectively). However, no differences emerged for scale condition at either Trial 1, $p = .128$, or Trial 2, $p = .218$.

3.3. Discussion

Experiment 1 found evidence of repeated scale measurement effects on pain intensity but not pain tolerance. Compared to a traditional 0–10 scale, participants using a numerically shorter scale rated their pain on a subsequent test as worse. The same was not true for an alternative

0–10 scale that modified the value labels rather than the numeric length.

4. Experiment 2

Experiment 2 replicated the design of Experiment 1 with two key changes. First, we dropped the modified 0–10 scale condition (which did not produce different effects than the traditional 0–10 scale) and tested only the short 0–4 scale against the traditional 0–10 one. Second, we included several measures of potential psychological mechanisms that may have played a mechanistic role in the effects of scale use on pain perception. These are explained further in the Method section. Tests of the scale conditions on pain intensity and tolerance were confirmatory, following from Experiment 1, but tests of the potential mechanisms were exploratory (we had no strong reason to predict one mechanism over another).

4.1. Method

4.1.1. Participants

We collected data from 174 undergraduate students (113 female, $M_{\text{age}} = 18.7$), who received course credit for participating in a 2-cell between-subjects experiment.

4.1.2. Procedure

Experiment 2 followed the same procedure and used the same materials as in Experiment 1, though without the modified 0–10 scale condition. Additionally, following pain testing, participants completed three measures of potential mechanisms for the repeated scale measurement effects, in random order.

A first possibility for the effects observed in Experiment 1 is that the 0–4 scale may have been more aversive to use than the 0–10 scale, because it gave participants relatively few numeric options for accurately expressing their level of pain. To examine this, we measured affective evaluation in both trials by asking participants, “How did you feel about having to rate your pain using this specific scale?” An image of the scale used was provided with each question, and responses were made on a -4 (Very negative) to $+4$ (Very positive) scale.

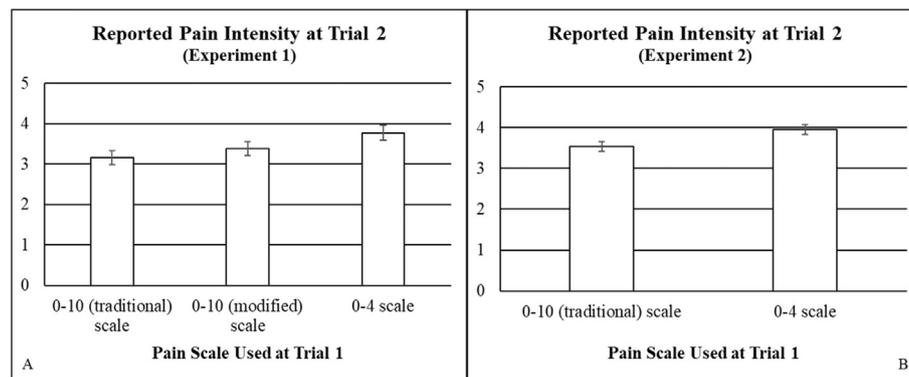


Fig. 3. Pain intensity (self-reported pain at removal of hand from cold pressor) at Trial 2 as a function of the scale used to report pain in Trial 1. Pain at Trial 2 was measured using a 0–6 scale. Experiment 1 (panel A) and Experiment 2 (panel B). Error bars reflect standard errors.

A second possibility is that the use of different scales affected the mental accessibility of different levels of past pain experiences. For instance, people using a 0–10 scale may recall worse prior pain events (those more indicative of a “10”) than people using a 0–4 scale. If so, one’s current pain may seem less bad when contrasted against very bad past experiences, leading to lower judgments of current pain. To examine this, we asked participants to think about past experiences they had in which they “felt pain at levels similar to or greater than what you felt in the cold water tasks” and to list as many of these experiences as quickly came to mind (maximum = 10). Along with each description, participants rated how painful that experience was on a 0 (not at all painful) to 6 (extremely painful) scale.

A final possibility we considered is that the Trial 1 scales could differentially affect the amount of pain that participants expected they could subsequently bear. For a given rating, such as pain that is 75% of the worst possible imaginable, the average response on shorter scales will necessarily be closer to the scale maximum (e.g., a 3 on a 0–4 scale) than the average response on longer scales (e.g., 7.5 on a 0–10 scale). Participants may be sensitive to the gap between their response and the scale maximum and interpret larger gaps as though they could “go further” in tolerating more pain. To examine this, two questions assessed expectations of additional tolerance for Trial 1 (“How much additional pain do you think you could have tolerated beyond what you felt during this task?”/“How much additional time do you think you could have kept your hand in the water beyond what you did during this task?”). Responses were made on 0 (could not have tolerated any more pain/kept my hand in any more time) to 8 (could have tolerated a lot more pain/kept my hand in a lot longer) scales. Before answering, participants were reminded of their actual pain rating from Trial 1.

4.2. Results

4.2.1. Data analysis plan

A total of 17 participants were excluded from all analyses for: failing to complete measures/being sick (5), relatively accurate suspicion (4), cold pressor task problems (1), and hand temperatures that did not return to within ± 5 °F of baseline before Trial 2 of the cold pressor (7). Additionally, 13 participants reported experiencing chronic pain, 7 more reported medication use that might affect pain sensitivity, and 6 more reported prior experience with cold pressor tasks. All were excluded from further analyses. We followed the same analytic procedures detailed in Experiment 1. After exclusions, our sample was sufficient to detect small-to-medium effects of $f \geq 0.247$ with 80% power in a between-subjects ANCOVA.

4.2.2. Primary analyses

For pain intensity, an ANCOVA on Trial 2 pain ratings revealed a main effect of Condition, $F(1, 127) = 5.77, p = .018, \eta_p = 0.04$.

Consistent with the findings in Experiment 1, patterns of the means indicated that the 0–4 scale led to higher pain ratings than the traditional 0–10 scale, $d = 0.41$ (95% CI: 0.074, 0.770). See Fig. 3 (Panel B).

For pain tolerance, time measurements on Trials 1 and 2 showed acceptable skewness ($< |1.51|$ and $< |0.97|$, respectively) and kurtosis ($< |1.24|$ and $< |0.17|$, respectively). As in Experiment 1, no differences between scale conditions emerged at either Trial 1, $p = .359$, or Trial 2, $p = .826$.

4.2.3. Potential mechanisms

All mechanism analyses were conducted controlling for the same factors as in the primary analyses. Examining the first potential mechanism— affective responses to the scales—a main effect of Condition emerged for Trial 1, $F(1, 127) = 10.67, p = .001, \eta_p = 0.08$, but not Trial 2, $F(1, 127) = 0.31, p = .582, \eta_p = 0.00$ (as one would expect given that the scale was identical across conditions in Trial 2). Consistent with this proposed rationale, participants rated use of the traditional 0–10 scale ($M = 7.03, SD = 1.70$) more positively than the 0–4 scale ($M = 5.92, SD = 2.09$). Controlling for this Trial 1 affective evaluation did not change the effect of the scale manipulation on pain intensity at Trial 2, however, suggesting that affect did not drive this effect.

Second, to evaluate whether the Trial 1 scales affected mental accessibility of past painful experiences, we calculated three scores: total number of recalled experiences, the average reported pain of those experiences, and the maximum reported pain across those experiences. A significant Condition effect emerged for maximum pain, $F(1, 127) = 5.26, p = .024, \eta_p = 0.04$, with participants who used the traditional 0–10 scale reporting a greater maximum intensity score for past pain experiences ($M = 6.32, SD = 0.83$) than participants who used the 0–4 scale ($M = 5.95, SD = 1.01$). This greater intensity in the 0–10 condition is consistent with the idea that participants may contrast current pain judgments against past pain judgments, thereby leading to less reported pain at Trial 2 in the 0–10 compared to the 0–4 condition. However, the evidence from these tests is weak. Further, controlling for the maximum intensity of past pain recalled did not change the effect of the scale manipulation on Trial 2 pain intensity (if anything, the repeated scale measurement effect was strengthened).

Finally, when examining expectancies that one could bear more of the cold pressor experience, a significant effect of Condition emerged for expected tolerance of additional pain, $F(1, 127) = 4.57, p = .035, \eta_p = 0.04$, and expected tolerance of spending additional time in the water, $F(1, 127) = 3.96, p = .049, \eta_p = 0.03$. The traditional 0–10 scale led to the perception that one could tolerate more pain ($M = 4.71, SD = 2.17$) and time in water ($M = 4.44, SD = 1.80$) compared to the 0–4 scale (pain: $M = 3.97, SD = 1.94$; time: $M = 3.78, SD = 1.77$). To examine these results further, we calculated a score for the difference between reported pain at Trial 1 and the condition-specific scale

maximum for that trial (as expected, participants in the 0–10 condition were much further from maximum than participants in the 0–4 condition, $F(1, 127) = 141.90, p < .001, \eta_p = 0.53$). We z-scored these differences and tested whether they correlated with expectations of additional tolerance. Distance from scale maximum positively correlated with expectations that one could tolerate additional pain, $r(127) = 0.33, p < .001$, and additional time in water, $r(127) = 0.33, p < .001$. These results are consistent with the idea that using a longer scale, and thus responding further from the scale maximum on average at Trial 1 (i.e., leaving a bigger gap between one's own response and the scale max), produces the inference that one can bear more pain in subsequent trials. Importantly, controlling for expectations of additional pain and time tolerance in the primary pain intensity analysis rendered the effect of scale condition non-significant, $F(1, 125) = 2.57, p = .111, \eta_p = 0.02$. We therefore conducted a bias-corrected test of simultaneous mediation, including all covariates, using the Hayes (2012) PROCESS procedure (Model 4, 10,000 bootstrap samples). We stress interpretive caution with this analysis as the mediators were measured after the dependent measures and thus causation cannot be inferred. The total effect of expectations did mediate the effect of scale condition on pain intensity (*indirect effect* = -0.155 ; 95% CI: $[-0.331, -0.014]$). Additionally, the specific expectations of additional pain tolerance also showed evidence of mediation when time tolerance was removed from the model (*indirect effect* = -0.131 ; 95% CI: $[-0.294, -0.005]$). Thus, the effect of the 0–4 scale on raising pain intensity relative to the traditional 0–10 scale may have been due, in part, to the smaller scale restricting participant beliefs that they could tolerate additional noxious experience.

5. General discussion

Can the very measures used to assess physical pain influence how people subjectively evaluate this experience? Two studies demonstrate that self-reports of pain intensity were worse when an earlier pain experience was judged on a shorter scale than on a longer scale. In Experiment 1, when participants used a 0–4 scale in Trial 1 (following the first cold pressor task), they perceived more pain in Trial 2 (the subsequent cold pressor) compared to participants who used either a traditional 0–10 scale or a 0–10 scale with modified value labels in Trial 1. In Experiment 2, this result replicated with a 0–4 scale versus a traditional 0–10 scale (we did not use the modified scale in Experiment 2).

The 0–4 scale appeared to produce a number of psychological consequences in comparison to the traditional 0–10 scale, from more negative affective responses to changes in the mental accessibility of past pain experiences. However, relative to the traditional 0–10 scale, changes in pain intensity as a result of the 0–4 scale appeared to be driven (statistically) only by the expectation that one could bear less additional painful experience.

As these studies represent an initial examination of carryover effects of pain scales, they include a number of limitations. We elicited pain through the cold pressor task, a commonly used pain stimulus, which can directly induce pain but also indirectly amplify pain responses through modulation of physiological processes such as inhibition of baroreceptor reflex function (e.g., Duschek et al., 2007; Suarez-Roca et al., 2019). Though possibly affecting interpretation of the pain response, this influence should be consistent across experimental conditions. Relatedly, we assessed pain with single-item measures. Although these are commonly recommended assessment tools (Fillingim et al., 2016), they tend to obscure or omit the multidimensional nature of pain (Williams et al., 2000). The numeric scales used at Trial 1 may better detect sensory-discriminative aspects of pain, whereas the faces scale used at Trial 2 may better detect affective-motivational aspects of pain. Despite this, we are aware of no theoretical reason to presume that carryover effects like those shown here are specific to one type of scale or dimension of pain. Finally, we recruited healthy undergraduate

participants in the present studies. This approach allowed us to focus on scale effects without accounting for the backgrounds and biases associated with individuals dealing with pain-related conditions (e.g., chronic pain sufferers), but it also limits our ability to generalize to such populations, who may respond in different ways to the same procedures.

This work has interesting implications for the measurement of experiences involving pain, and beyond. With regards to pain, our findings suggest that measurement tool decisions can have important consequences for how feedback is recorded and understood. Most directly, in contexts featuring multiple assessments of pain, the findings raise questions about the degree to which pain levels should be attributed to the choice of measurement (e.g., specific scales) versus other factors. Another interesting conclusion from these studies is that people appear more responsive to numeric variation in scales than to variation in scale labels. Changes in labels that should represent clear and important differences pain scoring (e.g., labeling 10 as “worst pain imaginable” vs. “somewhat bad pain” in Experiment 1) did not affect participant reports during either pain trial. If this relative insensitivity to label variation is reliable, it raises significant questions about how pain responses are understood in contexts such as health care. That is, health care providers may interpret self-reported pain scores in line with scale labels, whereas patients may ignore or discount labels in favor of whatever numeric values resonate with them (see also Williams et al., 2000).

We further might presume that some of the repeated scale measurement effects identified here extend beyond the domain of physical pain to other types of experiences, for instance anxiety and fatigue. These states are also often measured repeatedly when tracking the progress of interventions or the influence of time-varying predictors, and could therefore evince repeated scale measurement effects like those reported here. Of course, this is speculative and needs to be tested. One might suspect that the mechanism identified here (i.e., expectations about the amount of additional pain one could bear) limits the scope of experiences relevant to the scale effect we find; but, we would argue that expectations of additional tolerance are also relevant to states such as fatigue (e.g., Smets, Garssen, Bonke, & De Haes, 1995), anxiety (e.g., Beck, Epstein, Brown, & Steer, 1988), depression-happiness (e.g., McGreal & Joseph, 1993), or even more interpersonal constructs such as social avoidance and distress following rejection (e.g., Watson & Friend, 1969). In fact, a substantial body of literature demonstrates that psychologically powerful negative events such as social rejection can produce affective and sensory responses comparable to physical pain (e.g., Eisenberger, Lieberman, & Williams, 2003; Kross, Berman, Mischel, Smith, & Wager, 2011; MacDonald & Leary, 2005). Clearly, much more research remains to be done on the potential effects of scales in the context of repeated measurement.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jesp.2020.103961>.

Open practices

Open Materials and Open Data badges are requested for this manuscript. Materials and data for the experiments are available at https://osf.io/f9yvr/?view_only=d2aad4662d8349de81bb78aac060c720.

References

- Auvray, M., Myin, E., & Spence, C. (2010). The sensory-discriminative and affective-motivational aspects of pain. *Neuroscience & Biobehavioral Reviews*, *34*(2), 214–223.
- Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, *56*(6), 893–897.
- Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, *25*(3), 187–217.
- Brevik, E. K., & Skoglund, L. A. (1998). Comparison of present pain intensity assessments on horizontally and vertically oriented visual analogue scales. *Methods and Findings in Experimental and Clinical Pharmacology*, *20*(8), 719–724.

- Brinjikji, W., Luetmer, P. H., Comstock, B., Bresnahan, B. W., Chen, L. E., Deyo, R. A., ... Wald, J. T. (2015). Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. *American Journal of Neuroradiology*, 36(4), 811–816.
- Cleeland, C. S., & Ryan, K. M. (1994). Pain assessment: Global use of the Brief Pain Inventory. *Annals, Academy of Medicine, Singapore*, 23(2), 129–138.
- Coghill, R. C. (2010). Individual differences in the subjective experience of pain: New insights into mechanisms and models. *Headache: The Journal of Head and Face Pain*, 50(9), 1531–1535.
- Colloca, L., Sigauo, M., & Benedetti, F. (2008). The role of learning in placebo and placebo effects. *Pain*, 136(1–2), 211–218.
- Cruz-Almeida, Y., Riley, J. L., III, & Fillingim, R. B. (2013). Experimental pain phenotype profiles in a racially and ethnically diverse sample of healthy adults. *Pain Medicine*, 14(11), 1708–1718.
- Downie, W. W., Leatham, P. A., Rhind, V. M., Wright, V., Branco, J. A., & Anderson, J. A. (1978). Studies with pain rating scales. *Annals of the Rheumatic Diseases*, 37, 378–381.
- Duncan, L. A., Schaller, M., & Park, J. H. (2009). Perceived vulnerability to disease: Development and validation of a 15-item self-report instrument. *Personality and Individual Differences*, 47(6), 541–546.
- Duschek, S., Mück, I., & Del Paso, G. R. (2007). Relationship between baroreceptor cardiac reflex sensitivity and pain experience in normotensive individuals. *International Journal of Psychophysiology*, 65(3), 193–200.
- Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*, 302(5643), 290–292.
- Fillingim, R. B. (2005). Individual differences in pain responses. *Current Rheumatology Reports*, 7(5), 342–347.
- Fillingim, R. B., Loeser, J. D., Baron, R., & Edwards, R. R. (2016). Assessment of chronic pain: Domains, methods, and mechanisms. *The Journal of Pain*, 17(9), T10–T20.
- Foxen-Craft, E., & Dahlquist, L. M. (2017). Brief submaximal isometric exercise improves cold pressor pain tolerance. *Journal of Behavioral Medicine*, 40(5), 760–771.
- Garner, W. R. (1960). Rating scales, discriminability, and information transmission. *Psychological Review*, 67(6), 343–352.
- Goldberg, J., Weisenberg, M., Drobkin, S., Blittner, M., & Gotestam, K. G. (1997). Effects of manipulated cognitive and attributional set on pain tolerance. *Cognitive Therapy and Research*, 21(5), 525–534.
- Hartley, J., & Betts, L. R. (2010). Four layouts and a finding: The effects of changes in the order of the verbal labels and numerical values on Likert type scales. *International Journal of Social Research Methodology*, 13(1), 17–27.
- Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling. Retrieved from <http://www.afhayes.com/public/process2012.pdf>.
- Hicks, C. L., von Baeyer, C. L., Spafford, P. A., van Korlaar, I., & Goodenough, B. (2001). The Faces Pain Scale-Revised: Toward a common metric in pediatric pain measurement. *Pain*, 93(2), 173–183.
- Hoffman, G. A., Harrington, A., & Fields, H. L. (2005). Pain and the placebo: What we have learned. *Perspectives in Biology and Medicine*, 48(2), 248–265.
- Kaptchuk, T. J., & Miller, F. G. (2015). Placebo effects in medicine. *New England Journal of Medicine*, 373(1), 8–9.
- Kaptchuk, T. J., Stason, W. B., Davis, R. B., Legedza, A. R., Schnyer, R. N., Kerr, C. E., ... Goldman, R. H. (2006). Sham device v inert pill: Randomised controlled trial of two placebo treatments. *Bmj*, 332(7538), 391–397.
- Koenig, J., Jarczok, M. N., Ellis, R. J., Bach, C., Thayer, J. F., & Hillecke, T. K. (2014). Two-week test-retest stability of the cold pressor task procedure at two different temperatures as a measure of pain threshold and tolerance. *Pain Practice*, 14(3), E126–E135.
- Krebs, D., & Hoffmeyer-Zlotnik, J. H. P. (2010). Positive first or negative first? Effects of the order of answering categories on response behavior. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6(3), 118–127.
- Kross, E., Berman, M. G., Mischel, W., Smith, E. E., & Wager, T. D. (2011). Social rejection shares somatosensory representations with physical pain. *Proceedings of the National Academy of Sciences*, 108(15), 6270–6275.
- Leong, L. E., Cano, A., Wurm, L. H., Lumley, M. A., & Corley, A. M. (2015). A perspective-taking manipulation leads to greater empathy and less pain during the cold pressor task. *The Journal of Pain*, 16(11), 1176–1185.
- Leventhal, H., Brown, D., Shacham, S., & Engquist, G. (1979). Effects of preparatory information about sensations, threat of pain, and attention on cold pressor distress. *Journal of Personality and Social Psychology*, 37(5), 688–714.
- MacDonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological Bulletin*, 131(2), 202–223.
- Mazaheri, M., & Theuns, P. (2009). Effects of varying response formats on self-ratings of life-satisfaction. *Social Indicators Research*, 90(3), 381–395.
- McGreal, R., & Joseph, S. (1993). The depression-happiness scale. *Psychological Reports*, 73, 1279–1282.
- Modir, J. G., & Wallace, M. S. (2010). Human experimental pain models 2: The cold pressor model. In A. Szallasi (Ed.), *Analgesia* (pp. 165–168). New Jersey: Humana Press.
- Naugle, K. M., & Riley, J. L., 3rd (2014). Self-reported physical activity predicts pain inhibitory and facilitatory function. *Medicine and Science in Sports and Exercise*, 46(3), 622–629.
- Neville, S. J., Clauw, A. D., Moser, S. E., Urquhart, A. G., Clauw, D. J., Brummett, C. M., & Harte, S. E. (2018). Association between the 2011 Fibromyalgia Survey criteria and multisite pain sensitivity in knee osteoarthritis. *The Clinical Journal of Pain*, 34(10), 909–917.
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1–15.
- Racine, M., Tousignant-Laflamme, Y., Kloda, L. A., Dion, D., Dupuis, G., & Choinière, M. (2012). A systematic literature review of 10 years of research on sex/gender and experimental pain perception—part 1: Are there really differences between women and men? *Pain*, 153(3), 602–618.
- Rammstedt, B., & Krebs, D. (2007). Does response scale format affect the answering of personality scales? Assessing the Big Five dimensions of personality with different response scales in a dependent sample. *European Journal of Psychological Assessment*, 23(1), 32–38.
- Scheier, M. F., & Carver, C. S. (1985). The Self-Consciousness Scale: A revised version for use with general populations. *Journal of Applied Social Psychology*, 15(8), 687–699.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54(2), 93–105.
- Smets, E. M. A., Garssen, B., Bonke, B. D., & De Haes, J. C. J. M. (1995). The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research*, 39(3), 315–325.
- Snyder, M., & Gangestad, S. (1986). On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of Personality and Social Psychology*, 51(1), 125–139.
- Srinivas, S. V., Deyo, R. A., & Berger, Z. D. (2012). Application of “less is more” to low back pain. *Archives of Internal Medicine*, 172(13), 1016–1020.
- Suarez-Roca, H., Klingler, R. Y., Podgoreanu, M. V., Ji, R. R., Sigurdsson, M. I., Waldron, N., & Maixner, W. (2019). Contribution of baroreceptor function to pain perception and perioperative outcomes. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 130(4), 634–650.
- Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, 68(3), 368–393.
- Tourangeau, R., Couper, M. P., & Conrad, F. G. (2013). “Up means good” the effect of screen position on evaluative ratings in web surveys. *Public Opinion Quarterly*, 77, 69–88.
- Tracey, I. (2010). Getting the pain you expect: Mechanisms of placebo, nocebo and re-appraisal effects in humans. *Nature Medicine*, 16(11), 1277–1283.
- Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., ... Cohen, J. D. (2004). Placebo-induced changes in fMRI in the anticipation and experience of pain. *Science*, 303(5661), 1162–1167.
- Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, 33(4), 448–457.
- Williams, A. C. D. C., Davies, H. T. O., & Chadury, Y. (2000). Simple pain rating scales hide complex idiosyncratic meanings. *Pain*, 85(3), 457–463.
- Younger, J., McCue, R., & Mackey, S. (2009). Pain outcomes: A brief review of instruments and techniques. *Current Pain and Headache Reports*, 13(1), 39–43.