

PAI 01308

# The Descriptor Differential Scale: applying psychophysical principles to clinical pain assessment

Richard H. Gracely and Donna M. Kwilosz

*Clinical Pain Section, Neurobiology and Anesthesiology Branch, National Institute of Dental Research, National Institutes of Health,  
Bethesda, MD 20892 (U.S.A.)*

(Received 11 March 1988, revision received and accepted 29 June 1988)

---

**Summary** The Descriptor Differential Scale (DDS) applies psychophysical principles to clinical pain assessment. It contains 12 descriptor items for each pain dimension assessed. For each item, subjects indicate if their pain either is equal in magnitude to that implied by the anchoring descriptor, or how much greater or lesser on a 10-point graphic scale. The method permits collection of multiple responses, reducing scaling error, and assesses both pain magnitude and scaling consistency. Ninety-one patients completed the sensory intensity and unpleasantness forms of the DDS at both 1 and 2 h after surgical extraction of a lower third molar. Results show that the DDS satisfies standard psychometric criteria for reliability, objectivity and item homogeneity. The coefficients found satisfy standard psychometric criteria and improve after elimination of inconsistent profiles.

**Key words:** Descriptor Differential Scale; Pain

---

## Introduction

With the notable exception of the McGill Pain Questionnaire [11], most measures of clinical pain magnitude use a single measurement obtained from a visual-analog, numerical, or verbal scale of pain intensity. These simple scales are readily completed and easily scored, and have produced reliable data in many investigations of pain and pain control methods.

Despite their apparent utility, these scales treat pain as a simple unitary dimension, varying only in intensity. They also provide only one measurement per trial, potentially resulting in increased variability in comparison to psychophysical proce-

dures that collect 20–100 observations per trial. In addition, they rely on a spatial judgment or a specific response choice. Consequently they are vulnerable to several scaling biases, including the tendency to repeatedly use the same category or part of a line (reducing sensitivity in assessment of analgesics) or remembering a past discrete response (rather than the magnitude of previous pain in a study of pain memory).

This study presents data from the Descriptor Differential Scale (DDS), which is based on verbal descriptors used extensively for experimental pain assessment. Previous studies demonstrated the reliability, objectivity and validity of verbal scales of sensory intensity (e.g., mild, moderate, intense) and unpleasantness (e.g., annoying, unpleasant, distressing) quantified by ratio-scaling procedures [4,5]. Additional experiments using pain sensations produced by electrical stimulation of the skin or teeth or by heat applied to the skin have

---

*Correspondence to:* Dr. Richard H. Gracely, NAB, NIDR, NIH, Building 10, Room 3C403, 9000 Rockville Pike, Bethesda, MD 20892, U.S.A.

shown that these dimensions respond differentially to various manipulations, including administration of narcotic analgesics [2,3], information about expected sensations [12], or hypnotic suggestion [10].

Conventional psychophysical methods present multiple stimuli of varying intensities and collect multiple responses comparing the sensations evoked by these stimuli to an implicit or explicit subjective standard. These methods cannot be applied directly to the assessment of clinical pain, since there is only a single, uncontrolled ‘stimulus.’ The DDS, however, performs the converse operation. It presents multiple subjective standard stimuli and collects responses comparing these to the clinical pain stimulus. Like psychophysical procedures, it collects multiple responses that both minimize response variability and allow an evaluation of scaling consistency. The multiple descriptors cover the entire pain range, minimizing floor and ceiling effects. These features suggest that the DDS may be less sensitive to biases associated with the use of a single discrete or visual analog response. In addition, the evaluation of scaling consistency may identify non-complying patients or those who perform poorly in a scaling task. These patients can compromise experimental studies of pain and analgesia and may provide unreliable information in a clinical evaluation.

### Methods

Ninety-one dental patients ages 18–39 completed the sensory intensity and unpleasantness forms of the DDS at 1 and 2 h after surgical removal of a lower third molar under lidocaine with 1:100,000 epinephrine local anesthesia. The sensory intensity form is shown in Fig. 1. Each form presents 12 descriptors printed in random order by a computer. Each descriptor is centered over 21 horizontal dashes with a minus sign (‘-’) over the left dash and a plus sign (‘+’) over the right dash. Patients are instructed to rate the magnitude (sensory intensity, unpleasantness) of their clinical pain in relation to each descriptor. If their pain magnitude is equal to that implied by the descriptor, they check the dash directly below

EACH WORD REPRESENTS AN AMOUNT OF SENSATION.

RATE YOUR SENSATION IN RELATION TO EACH WORD WITH A CHECK MARK.

The form contains the following descriptors and checkmarks:

- FAINT: checkmark on the 18th dash
- MODERATE: checkmark on the 17th dash
- BARELY STRONG: checkmark on the 11th dash
- INTENSE: checkmark on the 13th dash
- WEAK: checkmark on the 19th dash
- STRONG: checkmark on the 15th dash
- VERY MILD: checkmark on the 18th dash
- EXTREMELY INTENSE: checkmark on the 12th dash
- VERY WEAK: checkmark on the 19th dash
- SLIGHTLY INTENSE: checkmark on the 13th dash
- VERY INTENSE: checkmark on the 13th dash
- MILD: checkmark on the 17th dash

Fig. 1. Sensory intensity DDS form. Twelve descriptor items are printed in random order by a computer. Each descriptor item is centered over 21 horizontal dashes with a minus sign (‘-’) over the left dash and plus sign (‘+’) over the right dash. Subjects indicate their pain in relation to each descriptor by checking a single dash in relation to the descriptor. A check on the dash under the descriptor indicates pain sensation equal to the intensity implied by the descriptor, a check on a dash to the left indicates pain intensity proportionally less than that implied by the descriptor, a check to the right indicates pain intensity proportionally more intense.

the descriptor. If it is greater than the descriptor, they check a dash to the right that indicates the degree to which their pain is greater. If their pain magnitude is less than that implied by the descriptor, they make an appropriate check to the left of the center dash. The result for each descriptor is a rating of pain magnitude on a 0–20 scale that indicates amount of pain in relation to that descriptor, with 10 indicating a magnitude equal to that implied by the word. The unpleasantness form is similar, with 12 randomly placed descriptors of unpleasantness (see Table 1) in place of sensory intensity descriptors.

## Results

### Reliability

Responses to each DDS form were recorded for each individual descriptor item and for a total mean score averaged over all 12 items. Figs. 2 and 3 compare the total scores for all patients obtained at hours 1 and 2 for the sensory intensity and unpleasantness forms. Means and standard errors were  $10.4 \pm 0.40$ ,  $10.3 \pm 0.37$  for sensory intensity and  $8.7 \pm 0.41$ ,  $9.1 \pm 0.38$  for unpleasantness. The Pearson product-moment correlations between hours 1 and 2 were 0.82 for sensory intensity and 0.78 for unpleasantness. Inspection of the figure shows a linear relation with outliers reporting minimal pain at hour 1, or showing a decrease in pain between hours 1 and 2. This

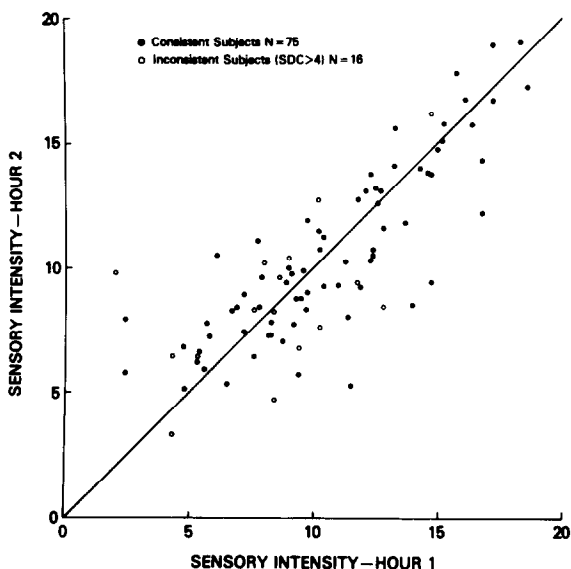


Fig. 2. Sensory intensity repeat reliability. Intensity scores obtained at hour 2 (mean 10.3, S.E. 0.37) are plotted against those at hour 1 (mean 10.4, S.E. 0.40) for 91 subjects. Scores above the line of unit slope increased, and those below decreased, over the hour. Repeat reliability, shown by the correlation between the scores, is 0.82. The open circles identify inconsistent individuals based on a comparison to group scores (see text and Figs. 5B, 6A and 7) or to themselves (see text and Figs. 5C, 6B and 8). A standard deviation-derived consistency measure (SDC) of 4 or greater was used to classify inconsistent subjects (see text and Figs. 7 and 8). Eliminating these subjects increased reliability to 0.84.

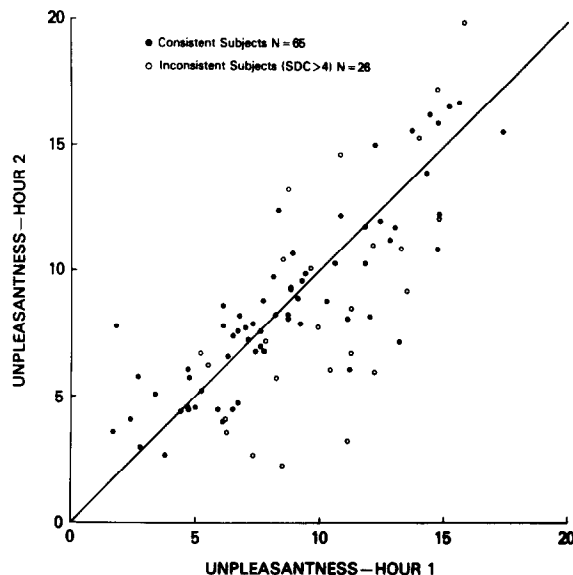


Fig. 3. Unpleasantness repeat reliability. Unpleasantness scores obtained at hour 2 (mean 9.1, S.E. 0.38) are plotted against those at hour 1 (mean 8.7, S.E. 0.41) for 91 subjects. Scores above the line of unit slope increased, and those below decreased, over the hour. Repeat reliability, shown by the correlation between the scores, is 0.78. The open circles identify individuals who scored 4 or greater on any SDC consistency measure. Eliminating these subjects increased reliability to 0.83.

measure provides an indication of repeat reliability.

### Item reliability and homogeneity

Table I shows Pearson product-moment correlations between hours 1 and 2 for individual items. These item reliability coefficients ranged from 0.61 to 0.83 (mean, 0.71) for the sensory intensity scale and 0.43 to 0.71 (mean, 0.59) for the unpleasantness scale. Item homogeneity can be expressed by correlating the scores on each item with the overall scale score computed from all other items [1]. These Pearson product-moment homogeneity coefficients also are shown in Table I. They range from 0.64 to 0.91 for sensory intensity and from 0.58 to 0.91 for unpleasantness. The sensory intensity items show an 'inverted-U' function with coefficients maximal for the middle descriptors and decreasing for extreme descriptors. This pattern is found at both hours 1 and 2 ( $R = 0.73$ ). The unpleasantness descriptors do not show this

TABLE I

## REPEAT-RELIABILITY COEFFICIENTS AND ITEM-HOMOGENEITY COEFFICIENTS FOR THE SENSORY-INTENSITY AND UNPLEASANTNESS FORMS OF THE DDS

The first column shows Pearson product-moment correlations between hours 1 and 2 for each individual descriptor item and for a total score for the whole form. The second column shows the same correlations for 'consistent profiles' (SDC scores less than 4 on the sensory intensity ( $n = 75$ ) or unpleasantness ( $n = 65$ ) form). Columns 3–6 show homogeneity coefficients for each item, defined as the Pearson product-moment correlation between the score on that item and an average score from the remaining items. Columns 3 and 5 show coefficients for all subjects for hours 1 and 2, columns 4 and 6 show the same correlations for 'consistent profiles' (SDC scores less than 4 on the sensory intensity ( $n = 75$ ) or unpleasantness ( $n = 65$ ) form). The repeat-reliability of these homogeneity coefficients is shown by Pearson product-moment correlations between hours 1 and 2.

Descriptor	Reliability		Homogeneity			
	Entire sample	Consistent subjects	Hour 1		Hour 2	
			Entire sample	Consistent subjects	Entire sample	Consistent subjects
<i>Sensory intensity</i>	N = 91	N = 75	N = 91	N = 75	N = 91	N = 75
Faint	0.72	0.72	0.70	0.78	0.66	0.83
Very weak	0.61	0.67	0.73	0.83	0.64	0.83
Weak	0.65	0.66	0.77	0.86	0.77	0.86
Very mild	0.68	0.77	0.81	0.89	0.73	0.84
Mild	0.77	0.78	0.88	0.92	0.84	0.90
Moderate	0.61	0.69	0.86	0.89	0.84	0.90
Barely strong	0.73	0.80	0.84	0.91	0.91	0.93
Slightly intense	0.83	0.84	0.81	0.89	0.84	0.90
Strong	0.79	0.79	0.81	0.90	0.83	0.92
Intense	0.81	0.79	0.80	0.91	0.82	0.89
Very intense	0.73	0.75	0.81	0.86	0.78	0.89
Extremely intense	0.63	0.77	0.67	0.80	0.73	0.84
Mean	0.71	0.75	0.79	0.87	0.78	0.88
Whole form	0.82	0.84	<i>R</i> between hours 1 and 2		0.73	0.84
<i>Unpleasantness</i>	N = 91	N = 65	N = 91	N = 65	N = 91	N = 65
Slightly unpleasant	0.66	0.70	0.70	0.77	0.74	0.82
Slightly annoying	0.59	0.67	0.78	0.85	0.82	0.85
Unpleasant	0.71	0.73	0.82	0.86	0.84	0.84
Annoying	0.58	0.73	0.79	0.80	0.79	0.83
Slightly distressing	0.60	0.72	0.74	0.82	0.83	0.89
Very unpleasant	0.64	0.76	0.84	0.88	0.91	0.90
Distressing	0.63	0.70	0.85	0.89	0.89	0.93
Very annoying	0.46	0.76	0.76	0.85	0.90	0.89
Slightly intolerable	0.57	0.75	0.74	0.89	0.75	0.87
Very distressing	0.68	0.78	0.84	0.90	0.89	0.93
Intolerable	0.54	0.74	0.66	0.87	0.82	0.87
Very intolerable	0.43	0.78	0.58	0.86	0.86	0.89
Mean	0.59	0.73	0.76	0.85	0.84	0.88
Whole form	0.78	0.83	<i>R</i> between hours 1 and 2		0.86	0.90

inverted-U pattern to the same degree as the sensory intensity descriptors. The unpleasantness coefficients also were similar between hours 1 and

2 ( $R = 0.86$ ). Mean homogeneity improved between hours 1 and 2 for the unpleasantness (0.76, 0.84) but not for the sensory intensity (0.79, 0.78) items.

### Scaling consistency

DDS scaling consistency can be assessed in several ways. Fig. 4 shows the responses in Fig. 1 rearranged by order of descriptor magnitude determined in a previous study by cross-modality ratio-scaling procedures [2]. Fig. 4 shows an actual result from a consistent subject. Item responses increase as descriptor magnitude decreases. Fig. 5A shows this subject's responses plotted against the log-ratio values determined for these descriptors in a previous study. The correlation ( $R = 0.97$ ) indicates a high degree of agreement between the DDS responses of this subject and a group scale determined by separate procedures with an independent group of subjects.

Fig. 5B shows the response made to each descriptor item for this same individual plotted against the mean responses of the rest of the group in this study. The linear relation ( $R = 0.98$ )

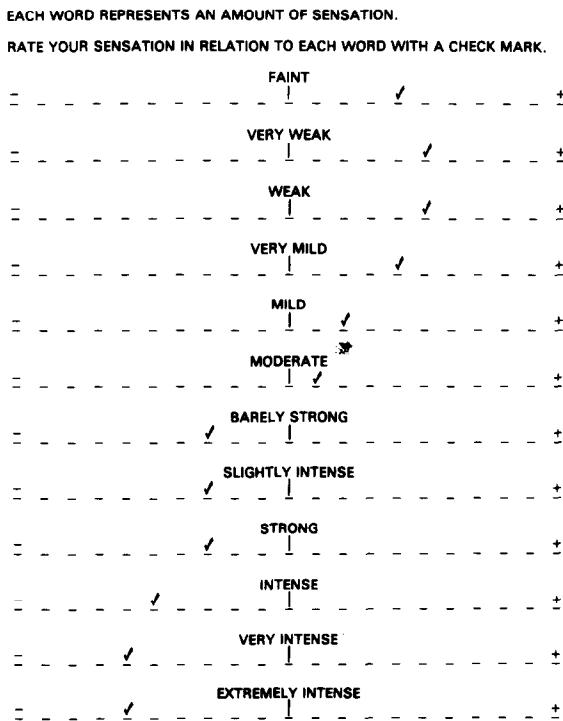


Fig. 4. Sensory intensity DDS form shown in Fig. 1 rearranged in order of increasing descriptor magnitude. Descriptor magnitude was determined in a previous study. Consistency is indicated by a monotonic relation; check marks move from left to right as descriptor magnitude decreases.

indicates a high degree of correspondence between this individual's responses and those of a similar group using the same response methodology. It also demonstrates a unique feature of this method. In addition to scaling pain magnitude, this procedure simultaneously determines the amount of intensity or unpleasantness implied by each descriptor. This property, described in detail in an unpublished report [9], allows comparisons of the word values produced during a pain rating with (1) word values determined separately by either the same or different procedures (Fig. 5A), (2) word values from a group in the same situation (Fig. 5B), or (3) word values determined previously by that individual (Fig. 5C). These comparisons check if subjects use the words in a manner consistent with their prior usage as well as with the group usage.

Fig. 5A and B compare the responses of a single illustrative individual to those of a group. This comparison results in a measure of 'to-group' consistency. In contrast, Fig. 5C shows a measure of 'individual' consistency. The same individual's responses, taken at hour 2, are plotted against that individual's responses observed at hour 1. The high correlation ( $R = 0.96$ ) between these responses indicates that this subject was consistent with his previous response choices.

The individual-consistency plot shown in Fig. 5C provides information about both consistency and analgesia. Analgesia is assessed by the decrease in vertical distance from the diagonal in this figure. The reduction in vertical distance indicates that for each individual item, the reported pain after an intervention (shown on the ordinate) is less than the pain reported before the intervention (shown on the abscissa). Analgesia also may be assessed simply as the reduction in total response following administration of a putative analgesic. Scaling consistency can be assessed by the correlation method described above or by other mathematical models. Kwilosz et al. [8] described a simple linear model that assesses consistency for each subject using an index equivalent to the standard deviation of the vertical distance from the diagonal for each descriptor data point. This standard deviation consistency (SDC) index is inversely related to consistency. Consistency

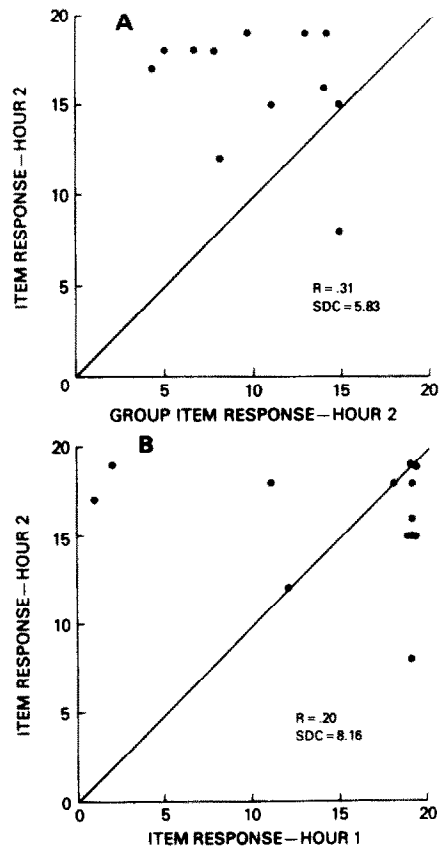
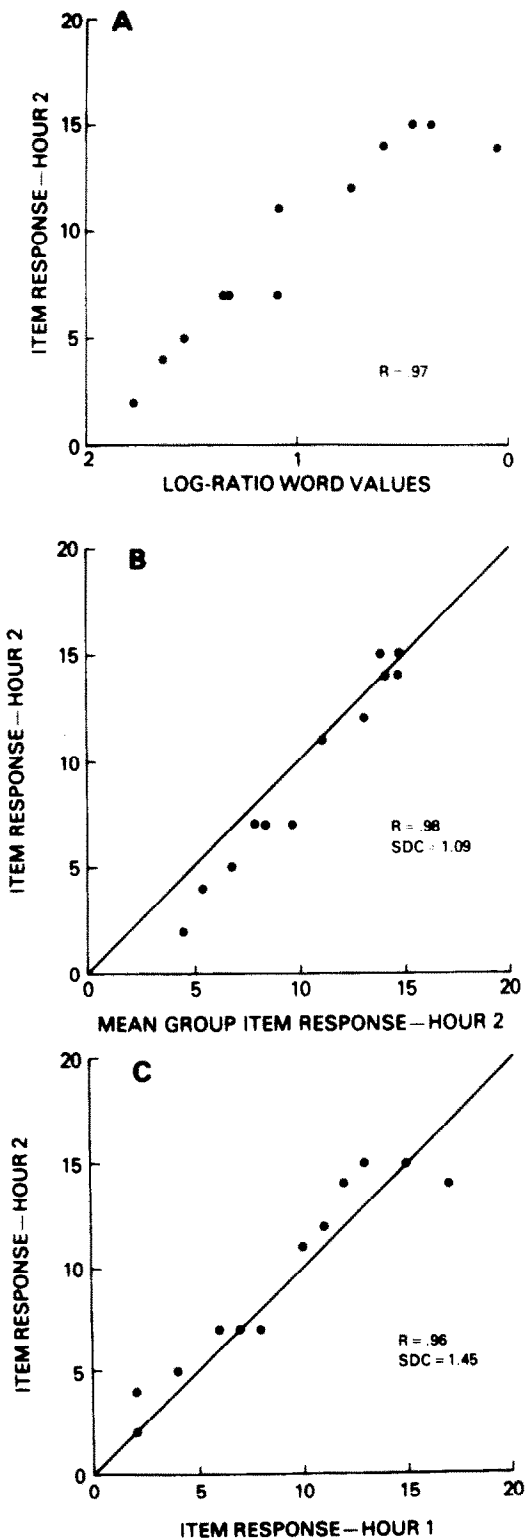


Fig. 6. Poor performance. Each point is the response to a specific sensory intensity descriptor identified in Table I. A shows the responses of an individual at hour 2 plotted against mean responses from the other 90 subjects. B shows responses of this individual at hour 2 against responses made at hour 1. Both plots show inconsistent responding, the points are non-linear.

decreases as the standard deviation increases; a standard deviation of 0 indicates data points parallel to the diagonal, representing perfect consistency. Using this model, analgesia is repre-

Fig. 5. Consistency measures in an individual subject. Each point is the response to a specific sensory intensity descriptor. A shows the responses shown in Fig. 4 plotted against log ratio-scaled values for each descriptor determined in a previous study. B shows the responses of the same individual plotted against the mean responses of the other 90 subjects. C shows these same responses, made at hour 2, plotted against the individual responses at hour 1. These plots show consistent responding. The relation is linear and parallel to the diagonal.

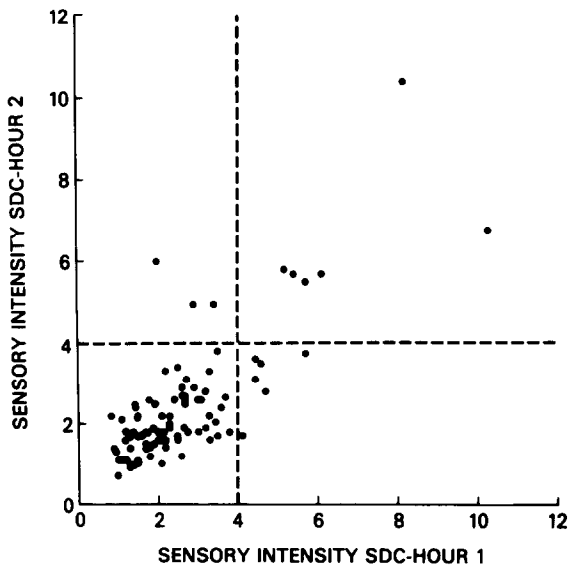


Fig. 7. Sensory intensity to-group SDC scores at hours 1 and 2. This SDC assesses inconsistency between individual responses and mean responses from the rest of the group (see Figs. 5B and 6A). Perfect consistency results in an SDC of 0. This figure shows SDC observed at hour 2 plotted against those found at hour 1 for each of the 91 subjects. The dashed lines show SDC cut-offs of 4.

sented by the mean of the differences from the diagonal, and performance consistency is represented by the standard deviation of these differences.

The 'to-group' and 'individual' standard deviation consistency measures (SDC) for Fig. 5B and C are 1.09 and 1.45, respectively. Fig. 6 shows to-group (A) and individual (B) SDC measures for another subject of 5.83 and 8.16. This subject used the descriptors in a manner inconsistent with both himself and a similar group.

Fig. 7 shows 'to-group' consistency scores for the sensory intensity form observed at hour 2 plotted against those found at hour 1. Mean SDC and standard errors were  $2.45 \pm 0.16$  and  $2.65 \pm 0.16$  for hours 1 and 2 respectively. Most subjects show high consistency (low SDC) as indicated by the bunching of the scores near the origin. A few subjects performed poorly during the first, second, or both hours. The present analysis uses an SDC criterion of 4 to classify inconsistent subjects. This criterion, shown by the dashed lines, separates the data points bunched near the origin from clear

outliers. Statistically, this criterion is greater than 2.5 standard deviation units from the mean to-group or individual SDCs for either the sensory intensity or unpleasantness scales. Thus these inconsistent cases represent less than 1% of the right tail of a normally distributed population. Inspection of individual response patterns of subjects with to-group SDC scores greater than 4 (see Fig. 6B) show that they could not or did not adequately perform the scaling task. Analysis of subjects with to-group SDCs less than 4 shows that SDC improved significantly ( $t(74) = 2.12$ ,  $P < 0.05$ ) between hours 1 and 2.

Unpleasantness scores also show a similar plot (hour 1 mean and standard error = 2.92, 0.15; hour 2, 2.39, 0.12). Performance of subjects with to-group SDCs less than 4 also improved between hours 1 and 2 ( $t(65) = 3.32$ ,  $P < 0.01$ ). Thus for both sensory intensity and unpleasantness forms, subjects became more consistent with a group usage between the first and second assessment times.

These results suggest that these scales can identify individual cases of poor consistency. Elimination of these cases would not bias outcome, since to-group SDC was uncorrelated with pain magnitude for both the sensory (hour 1,  $R = 0.09$ , 2,  $R = 0.11$ ) and unpleasantness (hour 1,  $R = 0.24$ ; 2,  $R = 0.12$ ) forms.

Fig. 8 shows individual SDC computed between hours 1 and 2 plotted against the to-group SDC at hour 2 for the sensory intensity form. This plot assesses item objectivity by comparing variance within an individual to variance between an individual and a similar group. This figure shows results similar to the to-group  $\times$  to-group plot shown in Fig. 7. Scores bunch at the low end of the unit diagonal and a cut-off of 4 identifies outlier subjects who did not perform consistently either with themselves or with a group norm. The equivalence of the individual (mean 2.51) and to-group (2.45) SDC scores indicates high item objectivity, since the similarity between individually determined and group-determined word scales is the same as the similarity of individually determined word scales over time. An individual's use of the verbal items is predicted a priori equally well by that individual or by the usage of a similar group.

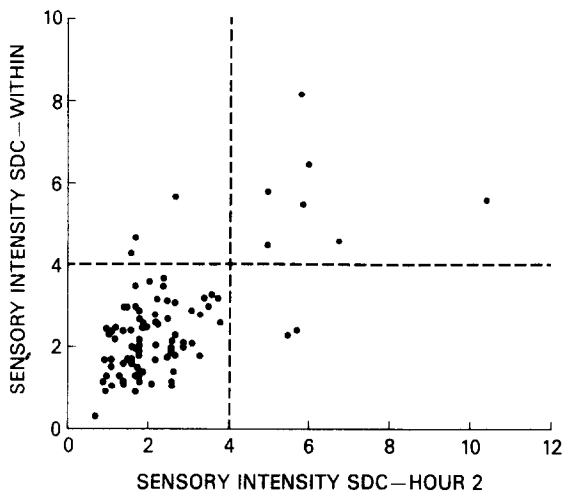


Fig. 8. Performance objectivity. SDC measures of performance are computed for each subject between hours 1 and 2 (individual SDC) and between hour 2 and a mean response from the rest of the group at hour 2 (to-group SDC). This figure shows within-SDC plotted against between-SDC for the sensory intensity scores from all subjects. It compares how well a subject's responses at hour 2 are predicted by his or her own responses at hour 1 versus how well they are predicted by a group norm. The dashed lines identify subjects with SDCs less than 4 that are consistent with both themselves and a group norm. The remaining subjects were inconsistent with either themselves, the group norm, or both.

Subjects with an individual or a to-group SDC of 4 or greater were classified as 'inconsistent.' Eliminating these subjects improved sensory intensity repeat reliability from 0.82 to 0.84 and unpleasantness repeat reliability from 0.78 to 0.83. These subjects are identified in Figs. 2 and 3. Many show a paradoxical decrease in pain between hours 1 and 2.

Table I shows the effect of inconsistent subjects on item reliability and homogeneity. Eliminating inconsistent subjects improved mean sensory item reliability from 0.71 to 0.75 and unpleasantness item reliability from 0.59 to 0.73. Item homogeneity improved for both the sensory (hour 1, 0.79–0.87; hour 2, 0.78–0.88) and the unpleasantness (hour 1, 0.76–0.85; hour 2, 0.84–0.88) forms.

## Discussion

These results show that the psychophysically based DDS satisfies standard psychometric

criteria. Pain intensity and unpleasantness were reliable over a 1 h period for both mean scores and for scores derived from individual items.

The obtained whole-form and item reliability coefficients can be influenced by several factors. Because of the 1 h retest interval, patients may remember and mark their previous responses, resulting in artifactually high reliability. This possibility is unlikely considering the nature of the task and the obtained item reliabilities. The scales are designed to reduce memory effects by presenting multiple descriptors in randomized sequence and requiring a 21-point rating scale response in relation to each descriptor. The ability to remember previous responses is assumed to be less than in scales in which specific items are chosen. This difficulty and the range of the reliability values shown in Table I strongly suggest that the duplication of previous responses was not a significant factor in this study.

The use of repeat correlations as reliability coefficients also assumes that pain levels varied considerably between patients and varied little within patients. The known variability of postoperative pain magnitude and the limited range of postoperative dental pain in comparison to more severe pain syndromes should lower repeat correlations. Thus, the obtained reliability is likely a conservative estimate, less than that expected with a population exhibiting a broad range of chronic pain magnitudes.

Item analyses further showed that each item was highly correlated with a mean scale derived from the other items, indicating that each scale is composed of a homogenous item pool. The sensory form showed a unique pattern of item scale correlations, maximal for descriptors of intermediate pain intensity and decreasing for more extreme intensities. This pattern suggests that the intermediate descriptors such as 'moderate' may provide a more representative measure than extreme descriptors such as 'faint' or 'extremely intense.' Alternatively, they may reflect floor and ceiling effects; the extreme dashes are checked more often for these descriptor items than for the less extreme items. Expanding the scale beyond 21 response possibilities may eliminate this effect.

The unpleasantness item-scale correlations did



not show the 'inverted U' pattern to the same degree as the sensory items. However, these correlations appeared to be specific to each descriptor, since the relative pattern did not vary over time. These stable patterns suggest that certain descriptors in the DDS may provide a better measure of the dimension assessed. Like other test instruments, overall homogeneity could be improved by an iterative process in which less homogeneous items are replaced by new items in a series of validation stages.

The DDS also produces independent measures of scaling consistency, a feature found in many psychometric instruments but not systematically applied to pain assessment. The structure of the DDS permits several types of consistency measures, including a correlation method and linear or non-linear variance models. The present analysis used a linear variance model [8] that assessed performance consistency either between an individual and a group norm or within an individual over time. The equivalence of these individual and to-group scores indicates that the magnitude associated with each descriptor item is similar within the group, supporting previous experimental findings [4]. These performance scores identified invalid profiles from subjects who either did not or could not scale their pain in a consistent manner. The correlation between the hour 1 and 2 consistency scores and the performance improvement over time suggest that scaling consistency may be a skill that improves with practice.

The reliability and homogeneity coefficients were improved by the elimination of invalid profiles. This improvement suggests that the inclusion of inconsistent subjects adds little and can possibly degrade the outcome of clinical trials relying on subjective pain reports. Similar to the use of validity scales in personality research, the identification and elimination of inconsistent subjects may significantly improve the power and efficiency of clinical pain studies. The assessment of the negative influence of inconsistent subjects and the identification of factors such as intelligence, personality, social or demographic variables that predict inconsistent subjects are logical goals of further research with this instrument.

In statistical terms, the present consistency

model and cut-off criterion minimize type I error, labeling subjects as inconsistent when they are not. All subjects who adequately scale their pain are included, only those who clearly do not consistently perform the scaling task are excluded. The model, as with most measures, is susceptible to type II error. It includes individuals with marginal consistency and may include individuals who produce responses internally consistent but not representative of their pain sensation. It identifies many, but not all, instances of poor performance.

The item homogeneity coefficients provide preliminary information about content validity. Additional preliminary evidence for content and construct validity is provided by another study of postoperative dental pain associated with third molar extractions. The DDS was used to assess pre-, intra-, and post-operative dental pain before and after double-blind intravenous administration of placebo, naloxone, fentanyl or diazepam to patients also receiving local anesthesia. Although no pain should be experienced under local anesthesia, patients reported a significant increase in pain magnitude after naloxone in comparison to placebo. Analysis of DDS responses showed a significant difference in the unpleasantness, but not the sensory intensity, forms [6,7].

The present form of the DDS and consistency analysis are preliminary steps in the development of an adequate pain assessment tool. The linear model may not adequately assess consistency in cases of faint or very intense pain, when many extreme categories are checked. The format may be difficult for some patients and requires comprehension of the verbal items. The DDS may not be useful as a single measure in all clinical situations. It may form part of an ideal test battery that assesses pain in stages that increase in both resolution and difficulty. Consistency measures at each stage would determine the responses that are accepted. All individuals would use the simplest scales, and those capable of using more sophisticated scales would be given an opportunity to do so.

In summary, these data suggest that the DDS is a reliable instrument that assesses pain magnitude and scaling behavior. It identifies inconsistent responding, and eliminating inconsistent profiles

improves its psychometric properties. Further studies are needed to assess the validity of the separate sensory intensity and unpleasantness scales, their sensitivity, and the influence of scaling consistency on the assessment of pain and analgesia.

### Acknowledgements

The authors gratefully thank Ron Dubner, Kenneth Hargreaves, James Lipton, Mitchell Max and Warren S. Torgerson for their critical comments and review of the manuscript.

### References

- 1 Cronbach, L.J., *Essentials of Psychological Testing*, 3rd Edition, Harper and Row, New York, 1970.
- 2 Gracely, R.H., Dubner, R. and McGrath, P.A., Narcotic analgesia: fentanyl reduces the intensity but not the unpleasantness of painful toothpulp sensations, *Science*, 203 (1979) 1261–1263.
- 3 Gracely, R.H., Dubner, R. and McGrath, P.A., Fentanyl reduces the intensity of painful toothpulp sensations: controlling for detection of active drugs, *Anesth. Analg.*, 61 (1982) 751–755.
- 4 Gracely, R.H., McGrath, P.A. and Dubner, R., Ratio scales of sensory and affective verbal pain descriptors, *Pain*, 5 (1978) 5–18.
- 5 Gracely, R.H., McGrath, P.A. and Dubner, R., Validity and sensitivity of ratio scales of sensory and affective verbal pain descriptors: manipulation of affect by diazepam, *Pain*, 5 (1978) 19–29.
- 6 Hargreaves, K.M., Neuroendocrine modulation of pain, *Anesth. Progr.*, 34 (1987) 121–123.
- 7 Hargreaves, K.M., Dionne, R.A., Mueller, G.P., Goldstein, D.S. and Dubner, R., Naloxone, fentanyl and diazepam modify plasma beta-endorphin levels during surgery, *Clin. Pharmacol. Ther.*, 40 (1986) 165–171.
- 8 Kwilosz, D.M., Gracely, R.H. and Torgerson, W.S., Memory for postsurgical dental pain, *Pain, Suppl.* 2 (1984) S426.
- 9 Kwilosz, D.M., Torgerson, W.S. and Gracely, R.H., Parallel forms of the DDS scale: theoretical and empirical equivalence, Unpublished manuscript.
- 10 Malone, M.D., The effects of hypnosis on verbal pain report. Unpublished doctoral dissertation, Department of Psychology, Washington University, St. Louis, MO, 1986.
- 11 Melzack, R., The McGill Pain Questionnaire: major properties and scoring methods, *Pain*, 1 (1975) 277–299.
- 12 Price, D.D., Barrell, J.J. and Gracely, R.H., A psychophysical analysis of experiential factors that selectively influence the affective dimension of pain, *Pain*, 8 (1980) 137–149.