Chem. Senses 31: 515–520, 2006

doi:10.1093/chemse/bjj054
Advance Access publication April 13, 2006

Validation of a Handheld Display Device for an Expandable Labeled Magnitude Scale (LMS)

Steve Guest[1], McSherry Wells[2], Chris Dancer[3] and Greg Essick[1]

[1]Center for Neurosensory Disorders, School of Dentistry, 2160 Old Dental Building, University of North Carolina at Chapel Hill, North Carolina, NC 27599-7450, USA, [2]Center for Neurosensory Disorders, School of Dentistry, University of North Carolina at Chapel Hill, North Carolina, NC 27599-7450, USA and [3]Dancer Design, 123 Boundary Road, St Helens, UK

Correspondence to be sent to: Steve Guest, School of Dentistry, 2160 Old Dental Building, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7450, USA. e-mail: steve_guest@dentistry.unc.edu

Abstract

Palmtop computers provide a possible avenue for the convenient collection of subjective ratings from individuals outside of a fixed laboratory setting. One disadvantage of these computers is the small size of their display screens, which may reduce the resolution of responses available as compared with standard display screens. One plausible solution to this problem is to use a scale that expands contingent on an initial response made by the subject, so that the final response is made from a scale with finer resolution. To validate this approach, we compared taste intensity judgments of six sucrose solutions (0.03, 0.06, 0.12, 0.24, 0.48, and 0.96 M), using a labeled magnitude scale either presented in expandable form on a palmtop computer (Palm scale) or in conventional (nonexpandable) form on a standard 17″ PC monitor (PC scale). Twenty-four subjects rated all six sucrose solutions thrice, using both scale types, the different scales being used on different days of testing. The scales led to very similar taste intensity ratings at all but the lowest concentration, which was rated less intense on the Palm scale. The Palm scale was used with slightly less precision than the PC scale for the weakest solution concentrations. In summary, the responses of the two scales were similar enough to validate the use of the expandable scale on the palmtop computer outside the laboratory setting.

Key words: labeled magnitude scale, palmtop computer, scale validation, taste

Introduction

In the early years of psychophysical research, the state of available technology meant that subjective ratings from individuals were obtained using pen and paper scales or verbal report. This changed with the advent of modern computers, and currently, responses are more commonly obtained using computers than any other method. In recent years, available input devices have developed to include technologies such as touch screens (Velikova *et al.*, 1999), which are a fundamental component of palmtop computers. Palmtop computers and indeed advanced mobile phones are capable of presenting response scales thus allowing the ready collection of data in the field with high compliance from experimental participants (Jamison *et al.*, 2001).

One problem with the use of handheld devices to present response scales is the relatively small size of their display screens. This is of particular consequence when using certain types of response scales, such as labeled magnitude scales (LMSs), because such scales typically have many descriptive intensity labels that need to be displayed, which is not practical on a small screen. Additionally, the popular LMS of Green *et al.* (1993) not only has several intensity labels but also three of the less intense labels are placed in the lower 17% of the scale. On a small screen this reduces the available resolution for responses made at the weak end of the scale. For example, the LMS displayed on a PC can easily have over 600 discrete steps (pixels) and perhaps as many as 1200, but on a small screen, it may be restricted to 100 steps. In the latter case, there are only 17 discrete values that may be selected between no sensation and a sensation of "moderate" intensity. Moreover, other scales such as the most recent incarnation of Borg's category-ratio (CR) scale for rating perceived exertion (E. Borg and G. Borg, 2002) are more complex still, with multiple labels, symbols, and icons that cannot all be represented on a small visual display. If these scales are to be used optimally on small display screens, then methods of addressing the above problem of limited resolution are necessary.

One possible way of dealing with complex response scales presented on small visual displays is to allow the scale to be zoomed or expanded, depending on an initial response obtained from the subject. If the initial response is made in a portion of the scale where the effective resolution is low, that section of the scale is then expanded to fill the screen, so improving the available resolution for a final response. To the best of our knowledge, no such scale has been suggested, tested, or reported in the literature. Here we compare the psychophysical characteristics of one such expandable LMS, displayed on a Palmtop computer, with the LMS displayed on a standard PC monitor. The questions of interest pertaining to an expandable versus nonexpandable scale format are as previously noted by Jamison *et al.* (2002) in the context of comparing paper- and computer-based scales. That is, do expandable and nonexpandable scales lead to the same ratings for a given stimulus, including ratings in the "expandable" portion of the scale, and can expandable and nonexpandable scales be used with equal precision?

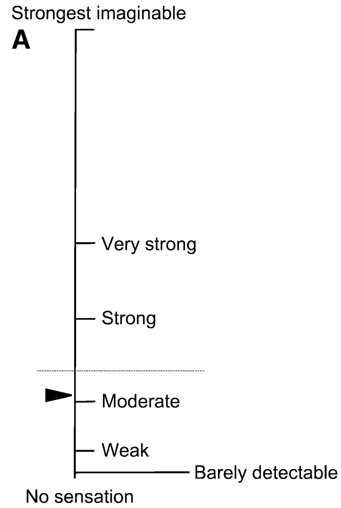

Figure 1 **(A)** Expandable (Palm) scale. Markers at each intensity label were 1 pixel in width and 4 pixels long, and the main scale line was 131 pixels in length. The dotted line shows the 25% point, the value below which the expanded scale **(B)**, consisting of 37% of the full scale, was shown. Note that this line was not shown to subjects in the experiment. The black triangle is the scale's pointer.

Materials and methods

Subjects

Twenty-four undergraduate and graduate students (19 male, 5 female, age range 19–24 years, median age 20 years) consented to take part in the study. All were compensated at a rate of $10/hour for their participation. The study was approved on ethical and safety grounds by the School of Dentistry Institutional Review Board at the University of North Carolina at Chapel Hill.

Stimuli and scales

The taste stimuli consisted of sugar (sucrose) solutions prepared in distilled water at six concentrations (0.03, 0.06, 0.12, 0.24, 0.48, and 0.96 M). These stimuli were presented at room temperature, in 10-ml samples, in individual 50-ml glass cups.

The scale used in the study was the LMS of Green *et al.* (1993), with one small change, namely the label "no sensation at all" was added below the lowermost end of the scale, in similar fashion to Borg's CR-100 scale (E. Borg and G. Borg, 2002). The LMS was displayed in vertical orientation on either a handheld display or a standard PC monitor. The handheld device was a Palm Zire m150 (Palm Inc., Sunnyvale, CA). The grayscale screen of this device is 4.6 cm^2 (160×160 pixels). The LMS (Palm scale) was drawn on the screen of this device such that it was 3.8 cm in length (131 pixels), which corresponds to about 11 degrees of visual angle when viewed from 20-cm distance. Responses on this scale were made using the standard Palm pen-style pointing device. The scale was programed so that any response below the 25% point of the scale (Figure 1A) would lead to the redisplay of the lower 37% of the scale (Figure 1B). If the initial response was below 25%, an intermediate screen was shown first that informed the subject that an expanded scale would be shown and that the response could be adjusted if the subject desired. As such, the scale from "no sensation" to "strong" was visible, redisplayed as 131 pixels (3.8 cm) in length (Figure 1B). When the expanded scale was shown, the cursor position was initially placed at the rating chosen by the subject using the unexpanded scale. The scale was generated by general purpose data collection software, running on the palmtop computer, and written by C. Dancer.

The PC scale was displayed on 17″ cathode ray tube (i.e., a conventional computer monitor). This scale was 600 pixels (18 cm) in length, corresponding to about 16 degrees of visual angle at a comfortable viewing distance of 60 cm. The scale was drawn and data were collected from it using Labview software (National Instruments, Austin, TX). This nonexpandable scale consisted of 600 discrete steps, and the subject made only a single response. The Palm and PC scales were not designed to subtend equal visual angles when in use. Instead, the scales were presented as they typically would be in an experiment, that is, using the maximum available screen space in both cases. The basic resolutions of the scales were also not the same. Instead, both were presented so that they used the maximum available resolution (i.e., 131 pixels for the Palm scale and 600 for the PC scale). As such, both scales appeared as continuous as possible when in use.

Procedure

The procedure used by Green *et al.* (1996) during validation of the LMS was adhered to closely. Each subject took part in two testing sessions on separate days. In one session, the subject made ratings with the PC scale, and in the other session, the Palm scale was used. Before commencing the ratings, subjects were given a short overview, lasting approximately 2 min, of the operation of the rating software (Palm or PC). The presentation of scales was randomized so that half of the subjects used the PC scale first, and half the subjects used the Palm scale first. In either case, it was emphasized to subjects that the upper intensity label of the whole scale referred to the strongest intensity of taste they could imagine, exclusive of painful taste sensations.

The stimuli were presented in random order, and between presentations of stimuli, the subject rinsed his or her mouth for 5 s with 10 ml of distilled water. All stimuli and rinses were expectorated. Each stimulus was rated three times within a session to give 18 ratings. After making each rating, the response scale was cleared with the pointer set to halfway along the scale so that the previous rating was not available when rating the next stimulus.

Statistical analysis

The two scales had different maximal resolutions, that is, 6 steps per percent of the PC scale, 1.31 steps per percent of the unexpanded Palm scale, and 3.51 steps per percent of the expanded portion of the Palm scale. Therefore, ratings from both scales were divided into 351 discrete values, which represents the maximum resolution of the Palm scale in its expanded state. This was carried out because the PC scale had greater resolution than the equivalent portion of the Palm scale, and such a difference could accentuate the disparity between responses using the two scales, especially with respect to response variability. In particular, one might predict more variable responses when using the PC scale for this reason alone. Subsequently, the log was taken of each rating, as is conventional when dealing with data obtained from an LMS (Green *et al.*, 1996). The ratings were then submitted to three analyses: first, a mixed-model analysis of variance (ANOVA) to determine the presence of any overall differences in ratings between the two scale formats; second, within-subjects correlations to determine agreement between scale ratings at the single subject level; and third, the analysis of the precision of responses made with each of the two scales.

Results

One subject's data were eliminated because that subject produced ratings on the Palm scale that were unrelated to the solution concentration, leaving 23 subjects with data available for analysis. A histogram of all ratings, prior to taking logarithms, is shown in Figure 2. One point of interest in this figure is the relatively high frequencies of responses in scale regions that include the label positions. This suggests a tendency in subjects to respond at (or near to) label positions, as opposed to in the between-label regions.

Agreement of ratings between scale formats

The means of the log ratings (±1 SE, $n = 23$) for the different concentrations as obtained from Palm and PC scales are shown in Figure 3. The data were first averaged over the three repeated ratings of each solution for each subject. Inspection of the figure suggests that, overall, any differences in ratings obtained from the two scale formats used in the

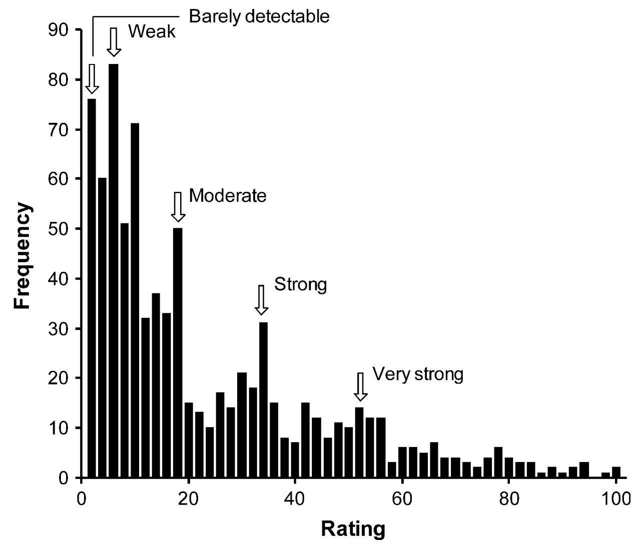

Figure 2 Histogram of all ratings, grouped into two unit wide bins. Arrows indicate the bars that included the rating value observed at each of the intensity labels.

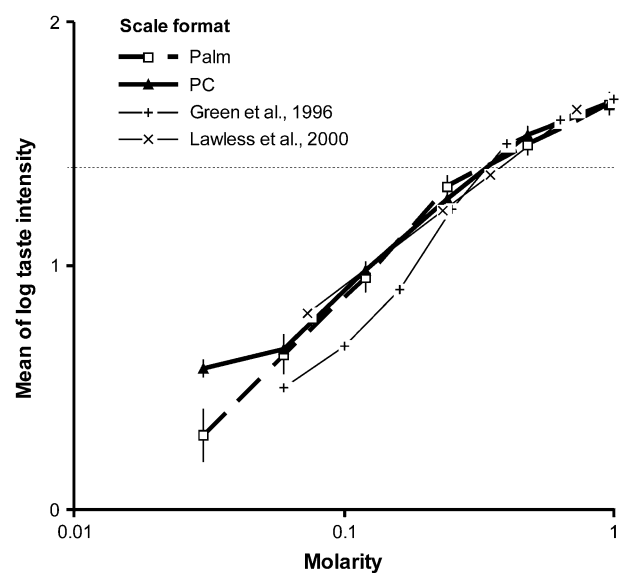

Figure 3 Taste intensity ratings given to six sucrose solutions when rated using an expandable LMS shown on a Palmtop computer or a conventional PC scale. The dashed horizontal line shows the break between where the Palm scale expanded (below the line) or did not (above the line). Error bars show ±1 SE. Sucrose intensity ratings from Green *et al.* (1996), Experiment Three, and Lawless *et al.* (2000), Experiment Three, are shown also.

current study were minimal, with the possible exception of the ratings given to the weakest (0.03 M) sucrose solution where the Palm scale led to lower ratings than the PC scale. Comparison with the studies of Green *et al.* (1996) and Lawless *et al.* (2000), who presented their scales on a PC, shows generally excellent agreement except that the Green *et al.* study found slightly lower ratings for concentrations in the range 0.1–0.16 M than found in the present study.

A scale format × repetition within session × order of scale use × concentration mixed-model ANOVA, carried out on the log-transformed raw data, confirmed that ratings varied according to sucrose concentration ($F_{5,735}$ = 375.8, P < 0.0001) but differently for the two scale formats ($F_{5,735}$ = 4.82, P < 0.001). However, pairwise (least square mean) tests between ratings obtained from the two scale formats at each concentration indicated that only the difference for the 0.03 M solution was statistically significant (P < 0.05). The 0.03 M solution was rated less intense with the Palm scale. An additional interaction with the scale format factor was identified, that of scale format × order of scale use ($F_{1,735}$ = 4.70, P = 0.031). This is illustrated in Figure 4. The order in which each scale was used changed the overall magnitude of ratings. Specifically, ratings from the Palm scale when it was used first were greater than those from the PC scale when used first. The reverse was true when comparing the scales on the second day. However, the largest mean difference between scales according to order of testing was just 0.17 units, representing 1.5% of the scale, and of little practical concern. Also of little importance was an interaction between repetition within session and concentration ($F_{10,735}$ = 2.48, P = 0.006). Ratings were slightly higher for the 0.06 M solution when it was rated the third time within a session.

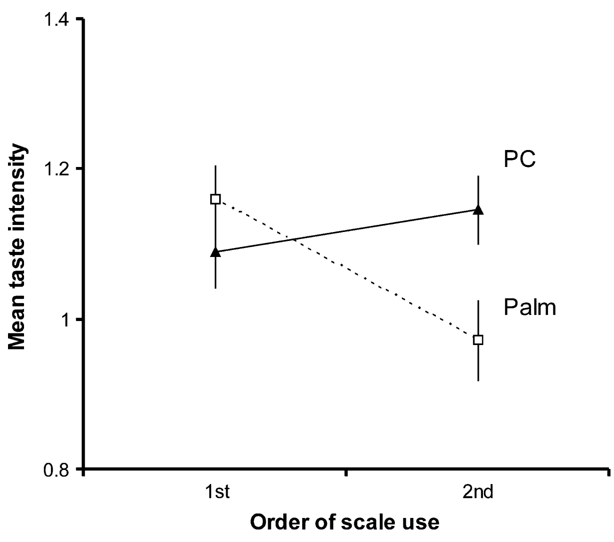

Figure 4 LMS taste intensity ratings averaged over six sucrose concentrations, plotted by session order in which a scale was used, for an expandable Palm scale and a conventional PC-based scale. The left-hand side of the figure indicates the ratings made when using Palm or PC scales on the first day of testing, whereas the right-hand side of the figure indicates ratings made on the second day of testing. Error bars show ±1 SE.

Within-subjects correlations

The above analysis indicates that, overall, Palm and PC scales led to very similar intensity ratings. The analysis was then extended to the level of individual subjects by calculating correlation coefficients between Palm and PC scale ratings for each subject. In each case, the average was taken over the three repetitions of each stimulus, for each scale, so that every coefficient was obtained from six data pairs. Correlations ranged from 0.73 to 1 (distribution shown in Figure 5), indicating that responses were indeed similar for Palm and PC scales, at the level of individual subjects.

Within- and between-subjects variability

The precision with which subjects made their ratings over multiple presentations of a given solution concentration was first analyzed. For each subject, the variance of the three log-transformed responses given for each concentration, for the two scale formats, was calculated. These within-subjects variances of logs were then entered into a mixed-model ANOVA (i.e., ANOVA of variance see Lederman and Taylor, 1972) with factors of scale format, order of testing, and concentration. The analysis indicated that, overall, the PC scale was used more precisely than the Palm scale ($F_{1,231}$ = 5.55, P = 0.019). Figure 6 illustrates the mean variance of logs plotted against the mean of the log ratings for each of the concentrations, for the two scale formats. From inspection of the figure, it appears that the difference in precision of use between the scale formats was only present for the lowest two concentrations (0.03 and 0.06 M). However, the interaction between scale format and concentration fell just outside significance ($F_{5,231}$ = 1.96, P = 0.084), thus failing to verify the trend toward greater Palm scale variability for the lower

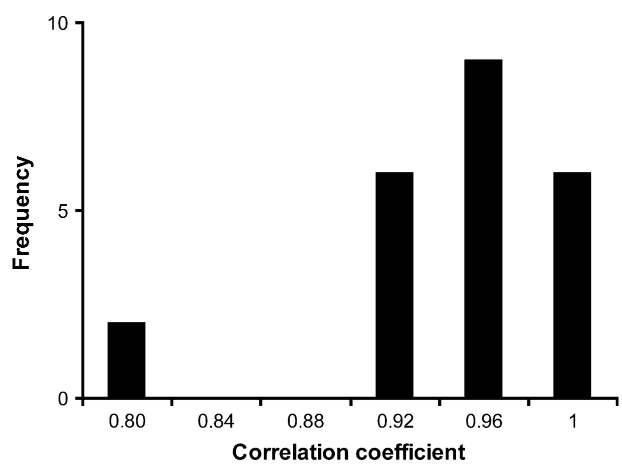

Figure 5 Distribution of the within-subjects correlations obtained between the average Palm and PC ratings for six sucrose concentrations. The abscissa labels indicate the upper boundary of each frequency bin.

concentrations. As is clear from Figure 6, in general, the precision increased (i.e., variability decreased) with increasing concentration ($F_{5,231}$ = 6.31, P < 0.0001).

In addition to the within-subjects variability analyzed above, between-subjects variability was considered by plotting the coefficient of variability (CV) for the two scale formats (Figure 7). Calculation of the CV used the formula for lognormally distributed data (Hastings and Peacock, 1975). The figure suggests that the between-subjects variability, when expressed as a proportion of the mean rating, was greater for the Palm scale. However, both scales were broadly similar in their CV values except for the Palm scale responses for the weakest two solutions.

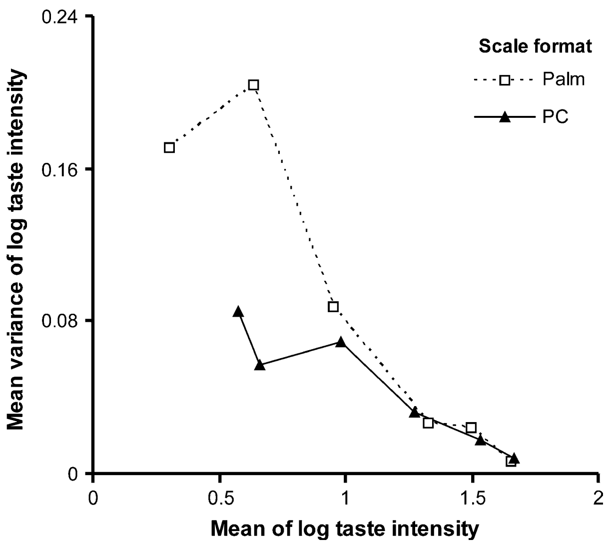

Figure 6 Mean within-subjects variance of logs for ratings made with Palm- or PC-based response scales, plotted against the mean (log) rating for each of the rated sucrose concentrations.

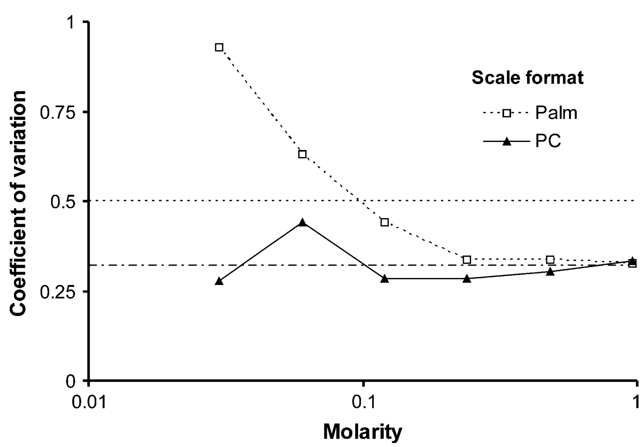

Figure 7 Coefficient of variation, expressing between-subjects variability, plotted for Palm- and PC-based response scales, for each of six rated sucrose concentrations. Dotted horizontal lines show the mean coefficient of variation for the two scale formats.

Discussion

Comparison of the intensity ratings of gustatory stimuli made with expandable Palm or conventional PC scales indicates that the two scale formats are generally similar in the results they lead to. This tolerance of response scales to different presentation systems is consistent with the results of Jamison *et al.* (2002), who found responses obtained from paper or palmtop computer presentation of visual analog scales were very similar. It was also the case that the sucrose intensity ratings in the experiment reported here agree reasonably well with sucrose intensity ratings obtained by others using the LMS displayed on a PC (Green *et al.*, 1996, Lawless *et al.*, 2000).

The one difference in the magnitude of ratings was for responses made in the very lowest portion of the Palm scale: the weakest (0.03 M) solution was rated less intense when the Palm scale was used as compared to the PC scale. Thus, the ratings obtained from the (expanded) Palm scale describe a steeper psychometric function than the ratings from the PC scale (see Figure 3, Palm scale points below the horizontal dashed line). Such a steepening is often seen as the stimulus range is reduced (Stevens, 1975) and has been shown to occur for the specific case of the LMS (Lawless *et al.*, 2000, Diamond and Lawless, 2001). The effect has been suggested as a consequence of a variety of possible contextual biases that may follow from stimulus range restriction, such as range equalization biases (Poulton, 1989). As such, one might propose any differences in ratings between Palm and PC scales to be a consequence of stimulus range restriction for the expanded Palm scale. However, in the case of the experiment reported here, the overall stimulus range was not specifically restricted and thus range-restriction effects would seem to require that the expanded Palm scale be treated as a separate entity with its own narrowed stimulus context. This may be unlikely given that recent experience of stimuli has been shown to influence subsequent scaling results (Stevenson and Prescott, 1994).

However, if subjects failed to maintain the appropriate, implicit, top anchor when viewing the expanded Palm scale, this might explain certain changes in the ratings made. For example, if the "strongest sensation imaginable" was not maintained as an implicit anchor, the expanded scale may have effectively become a new scale, with a new top anchor. If the new top anchor was interpreted as less intense than the anchor in the unexpanded scale, then subjects would have shifted their responses toward the more intense end of the scale as a consequence. That the intermediate intensity labels are judged with reference to the top anchor of the scale was shown by Green *et al.* (1996); if the top anchor of the LMS is changed from referring to all oral sensations, inclusive of pain, to all oral sensations, exclusive of pain, the slope of the relationship between perceived intensity and sucrose concentration changes, consistent with all intensity labels being judged in this new context (Green *et al.*, 1996).

In the expanded Palm scale data, the direction of the observed shift in the 0.03 M responses (i.e., toward less intense ratings) would require the top anchor in the expanded scale (strong) to be more intense than in the unexpanded scale. This seems a plausible change, given that the label "strong" was shown in the context of even stronger sensations ("very strong" and "strongest imaginable") in the unexpanded scale, whereas in the expanded scale, strong was the most intense label available. However, if this context-related effect did indeed apply, all ratings made using the expanded scale would be expected to shift toward lower intensities. This did not occur.

An alternative possibility for the shift is that expanding the scale facilitated making responses toward the weak end of the rated sensations. For example, in the expanded scale, the amount of screen space dedicated between a "barely detectable" and a "weak" sensation is far greater than in the unexpanded scale. This improved use of screen space might better allow subjects to make responses in the lower regions of the scale. That the separation in mean ratings between the 0.03 and 0.06 M solutions was greater on the Palm scale than the PC scale (see Figure 3) suggests potentially greater sensitivity for the Palm scale for rating very weak sensations.

In addition to the similarities between ratings obtained from Palm and PC scales, the precision of responses were broadly similar between scale formats. This was also true for between-subjects measures of variability, although the CVs were slightly higher for the Palm scale. Most of the greater variability in using the Palm scale was evident in responses to the weakest two solutions, although not the 0.12 M solution; all three of these solutions were, on average, rated using the expanded portion of the Palm scale. The generally slightly larger variability when using the Palm scale was probably due to the relative lack of familiarity of the subjects with the Palm input system itself. Nevertheless, it seems that the expandable scale shown on a small screen was almost as consistently and precisely used as standard PC monitor–based response system, and the mean ratings for the different test concentrations were very similar indeed, regardless of the scale format used. Thus, in conclusion, we suggest that zooming or expansion of an LMS is an acceptable way of implementing the scale on a small input device.

One final point of interest in the data was a mild tendency of subjects to respond at the label positions; these positions attracted subjects' responses disproportionately often. This tendency represents a potential problem to be borne in mind when choosing to use an LMS or similar scale and an issue that could reduce the sensitivity of the scale. It is of course possible that each of the labels was an excellent, near-optimal descriptor of one of the various sucrose concentrations, although this seems unlikely.

A download of the Palm software is available at http://www.dancerdesign.co.uk/.

Acknowledgements

We acknowledge the contribution of Dr Barry Green of the John B. Pierce laboratory and Yale University, who invented the original concept of an expandable LMS.

References

Borg, E. and **Borg, G.** (2002) *A comparison of AME and CR100 for scaling perceived exertion*. Acta Psychol. (Amst.), 109, 157–175.

Diamond, J. and **Lawless, H.T.** (2001) *Context effects and reference standards with magnitude estimation and the labeled magnitude scale*. J. Sens. Stud., 16, 1–10.

Green, B.G., Dalton, P., Cowart, B., Shaffer, G., Rankin, K. and **Higgins, J.** (1996) *Evaluating the 'labeled magnitude scale' for measuring sensations of taste and smell*. Chem. Senses, 21, 332–334.

Green, B.G., Shaffer, G. and **Gilmore, M.M.** (1993) *Derivation and evaluation of a sematic scale of oral sensation magnitude with apparent ratio properties*. Chem. Senses, 18, 683–702.

Hastings, N.A.J. and **Peacock, J.B.** (1975) Statistical Distributions. John Wiley & Sons, Inc., New York.

Jamison, R.N., Gracely, R.H., Raymond, S.A., Levine, J.G., Marino, B., Herrmann, T.J., Daly, M., Fram, D. and **Katz, N.P.** (2002) *Comparative study of electronic vs. paper VAS ratings: a randomized, crossover trial using healthy volunteers*. Pain, 99, 341–347.

Jamison, R.N., Raymond, S.A., Levine, J.G., Slawsby, E.A., Nedeljkovic, S.S. and **Katz, N.P.** (2001) *Electronic diaries for monitoring chronic pain: 1-year validation study.* Pain, 91, 277–285.

Lawless, H.T., Horne, J. and **Spiers, W.** (2000) *Contrast and range effects for category, magnitude and labeled magnitude scale in judgements of sweetness intensity.* Chem. Senses, 25, 85–92.

Lederman, S.J. and **Taylor, M.M.** (1972) *Fingertip force, surface geometry and the perception of roughness by active touch*. Percept. Psychophys., 12, 401–408.

Poulton, E.C. (1989) *Bias in quantifying judgments*. Lawrence Erlbaum Associates, London.

Stevens, S.S. (1975) Psychophysics: Introduction to its Perceptual, Neural and Social Prospects. John Wiley & Sons, New York.

Stevenson, R.J. and **Prescott, J.** (1994) *The effects of prior experience with capsaicin on ratings of its burn*. Chem. Senses, 19, 651–656.

Velikova, G., Wright, E.P., Smith, A.B., Cull, A., Gould, A., Forman, D., Perren, T., Stead, M., Brown, J. and **Selby, P.J.** (1999) *Automated collection of quality-of-life data: a comparison of paper and computer touch-screen questionnaires*. J. Clin. Oncol., 17, 998–1007.

Accepted March 22, 2006