R02 - 30

Appetite (2002) 39, 181–183
doi:10.1006/appe.2002.0496, available online at http://www.idealibrary.com on IDEAL®

Rapid Communication

A comparison of satiety measures

E. P. Merrill, F. M. Kramer, A. Cardello and H. Schutz

Behavioral Sciences Division, U.S. Army Natick Soldier Systems Center, Natick, MA 01670-5020, U.S.A

(Published electronically 5 August 2002)

Studies of human eating behavior have commonly used self-report scales to assess perceived appetite, hunger, and fullness (Burley *et al.*, 1987; De Graaf, 1997; Green & Blundell, 1996; Rolls *et al.*, 1990). This technique requires subjects to assign numerical ratings to index the magnitude of their subjective sensations (hunger, satiety, desire/motivation to eat, etc.). Subjective ratings are often assessed over time to obtain temporal profiles of these sensations. Such scales have also been considered to be valid indices of the strength of the individuals' appetite (Green *et al.*, 1997; De Graaf, 1993) and to correlate with food intake (Teghtsoonian *et al.*, 1991). While there are many scales for the assessment of subjective sensations, few studies have compared them to one another. The purpose of this study was to assess the sensitivity and reliability of five different 100 mm visual analogue scales of satiety. The scales consisted of (1) a bipolar hunger-fullness scale, with end-points labeled at bottom and top with "Extremely Hungry" and "Extremely Full" (2) a unipolar hunger scale, labeled at bottom and top with "Extremely Hungry" and "Not at all Hungry" (3) a unipolar fullness scale, labeled "Not at all Full" and "Extremely Full" (4) a unipolar "amount-could-eat" scale, labeled "A large amount" and "None at all" and (5) a 7-pt equal interval, bipolar scale of hunger/fullness, labeled from bottom to top with "Extremely Hungry", "Hungry", "Semi-Hungry", "No particular feeling", "Semi Satisfied", "Satisfied", and "Extremely Full" (Holt *et al.*, 1995). In this study, Scale 5 was treated as a VAS instead of a category scale in order to be consistent with the other scales used.

Study volunteers were employees of Natick Soldier Systems Center, most of whom had participated in previous food acceptability tests in the Center's consumer laboratory. Subjects were asked to complete a background questionnaire to determine morning eating routines. A total of 19 subjects participated in testing. Most participants were within the healthy weight range for their height (mean BMI ± SD, 26 ± 3·6 lbs). Foods used in the study were pilot tested to assess their acceptability and perceived satiety. All subjects reported that they found the test foods acceptable before commencing the study. Based on the pilot study and previous satiety literature (Holt *et al.*, 1995) four foods expected to vary in satiety value were selected. The foods used were strawberry yogurt (Axelrod), multigrain bread (Arnold), maple & brown sugar oatmeal (Quaker Instant), and croissants (Pillsbury Crescent). All foods were prepared according to manufacturer directions in 240 kcal portions on the morning of the test and presented in standard serving dishes.

A within-subjects, repeated measures design was used. Subjects were instructed not to eat after 9:00 pm on the evening before each test session to establish an overnight fasting time of 10–12 h. On each of four consecutive mornings, subjects reported to the research dining area at a scheduled time closest to their usual breakfast time (7:15–9:45 am). Upon arrival, subjects were asked to rate their baseline subjective hunger/fullness on each of the 5 test scales. Participants were instructed to pay attention to the wording of each scale since they could be similar. Immediately after completing these baseline ratings, subjects were presented one of the test foods for breakfast. No beverage was provided. Subjects were allowed to eat the food at a comfortable rate but were asked to finish within 10 min.

Immediately after consuming the food, subjects rated their liking/disliking of it using a nine point hedonic scale (Peryam & Pilgram, 1957). Following this and at every 10-min interval for the next 60 min, subjects rated their subjective hunger/fullness using the 5 test scales presented in random order. Subjects were allowed to read, relax, or talk with one another between ratings, with the exception that they could not discuss the food or the study in any way. At the end of each session, an *ad libitum* continental breakfast of cereal, muffins, bagels, tea, coffee, orange juice, and milk was provided. Each test session lasted approximately 1 h and 15 min. The same subjects returned one week later, and the procedure was repeated with the same test foods in the same order to assess test-retest reliability.

A repeated measures analysis of variance (ANOVA) was conducted to examine the relationship between satiety scales, foods, time, and session. The inverse of ratings was used for Scale 4 and the baseline rating was subtracted from all ratings to create a difference score. An item analysis among scales was used to measure the relationship between scales. In addition,

0195 6663 02/$ see front matter

the overall satiety response to a food was indexed by the area under the 60-min response curve (AUC). Area under the curve was calculated using the trapezoidal rule. An ANOVA of the average area under the curve was used to assess differences in satiating capacity of the foods.

Changes in visual analog scale ratings of all sensations followed the same general pattern. Figure 1 shows the mean difference from baseline in fullness and hunger ratings over time for each scale. As expected, there was a main effect for time ($F = 102{\cdot}7$, df = 1, 165, $p < 0{\cdot}001$), with all measures declining monotonically over time. In addition, there was a main effect of foods ($F = 14{\cdot}4$, df = 3, 365, $p < 0{\cdot}001$) and a time by food interaction ($F = 3{\cdot}9$, df = 3, 365, $p < 0{\cdot}001$). The mean difference in satiety ratings over time was greatest for oatmeal, followed by bread, croissant, and yogurt. These findings are generally consistent with previous work by Holt *et al.* (1995), but Holt reported yogurt to be slightly more satiating than croissants.

Item total correlations among scales showed that the bi-polar hunger/fullness scale (Scale 1) had the highest correlation for both week 1 ($r = 0{\cdot}94$) and week 2 ($r = 0{\cdot}93$). This was followed in order by the Holt bi-polar hunger/fullness scale (Scale 5) ($r = 0{\cdot}88$, $r = 0{\cdot}83$), the unipolar fullness scale (Scale 3) ($r = 0{\cdot}88$, $r = 0{\cdot}85$), the unipolar hunger scale (Scale 2) ($r = 0{\cdot}86$, $r = 0{\cdot}86$), and and the "amount could eat" scale (Scale 4), which had the lowest correlation with other scales for both week 1 ($r = 0{\cdot}76$) and week 2 ($r = 0{\cdot}72$). Although, the generally high correlations show that all of the scales have a strong association with one another. ANOVA with *post hoc* tests showed differences in the sensitivity of the scales to differences in satiety among the foods. In particular, it was found that the number of significant mean differences between foods was greatest for the unipolar hunger scale (Scale 2) and lowest for the unipolar fullness scale (Scale 3). An ANOVA conducted on the average area under the curve also demonstrated a main effect of scale ($F = 5{\cdot}0$, df = 4, 349,

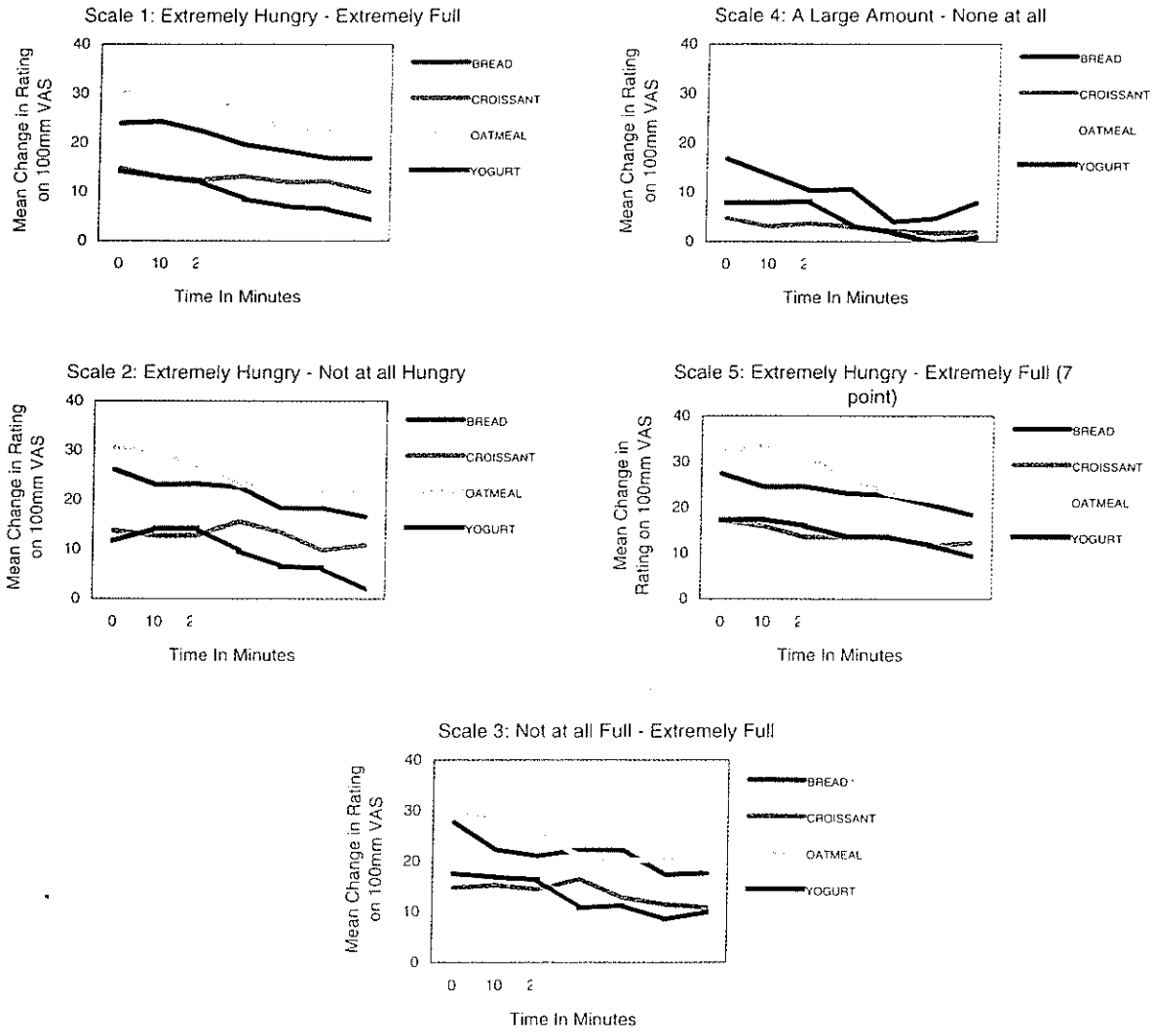

Figure 1. Mean change from baseline in fullness and hunger ratings over time for each scale.

$p \leq 0.001$) with multiple comparison tests showing Scale 4 to produce significantly smaller areas under the curve than all other scales.

Analysis of variance showed no main effect of replication for any scale. Pearson correlation coefficients between the ratings for week 1 and 2 showed that the Holt bi-polar hunger/fullness scale (Scale 5) had the highest correlation ($r = 0.64$), followed by the unipolar hunger scale (Scale 2, $r = 0.60$), the bi-polar hunger/fullness scale (Scale 1, $r = 0.50$), the unipolar fullness scale (Scale 3, $r = 0.44$), and the "amount could eat" scale (Scale 4, $r = 0.37$) which had the lowest correlation.

While it is not possible to conclude that any one of these scales is best for measuring perceived satiety, the data do indicate that reliability and sensitivity differences exist among the scales. For example, the unipolar "amount could eat" scale was found to be the least reliable and had the lowest association with the other scales. While, the two bi-polar hunger/fullness scales (Scale 1 and Scale 5) are very similar as far as the wording of the endpoints, Scale 1 (total number of differences = 11) had greater sensitivity than Scale 5 (total number of differences = 8) but Scale 5 had slightly higher reliability than Scale 1. Nonetheless, taking both sensitivity and reliability into consideration, the unipolar hunger scale (Scale 2) had the greatest sensitivity and high week to week reliability.

Another aspect of the data concerns the total amount of time required to index the satiety value of a food. Measuring satiety over the course of one hour may be underestimating the area under the curve for more satiating foods, such as oatmeal, because rated satiety levels remained above baseline level even after one hour. Therefore, a somewhat longer period of time may be necessary to accurately predict the satiating capacity of such foods. However, correlations among ratings at 30 and 60 min showed a high degree of association between the ratings at 30 and 60 min ($r = 0.88$), suggesting that a 30-min testing period may be as predictive of the satiety value of a given food as a 60-min testing period.

Perhaps one of the more interesting findings in this study is that participants were best able to reliably discriminate among foods through their ratings of hunger (Scale 2). When the sensitivity and reliability difference between Scale 2 and Scale 3 are considered, it becomes apparent that an individual may have an easier time differentiating feelings of hunger than feelings of fullness. This finding is an important one for future investigations of appetitive behavior.

It may well be the case that the use of these scales in combination with each other would provide the researcher with a better understanding of the subjective aspects of satiety. Future research will be aimed at developing an optimal tool for assessing satiety, including the development and testing of a labeled magnitude scale of satiety.

References

Burley, V. J., Leeds, A. R. & Blundell, J. E. (1987). The effect of high and low-fibre breakfasts on hunger, satiety, and food intake in a subsequent meal. *International Journal of Obesity* **1**, 87–93.

De Graaf, C., De Jong, L. L., Lamber, A. C. (1998). Palatability affects satiation but not satiety. *Physiology and Behavior* **4**, 681–688.

De Graaf, C. (1993). The validity of appetite ratings beliefs. *Appetite* **21**, 156–160.

Green, S. M. & Blundell, J. E. (1996). Subjective and objective indices of the satiating effects of foods: Can people predict how filling a food will be? *European Journal of Clinical Nutrition* **50**, 798–806.

Green, S. M., DeLargy, H. J., Joanes, D. & Blundell, J. E. (1997). A satiety quotient: A formulation to assess the satiating effect of food. *Appetite* **29**, 291–304.

Holt, S. A., Miller, J. C., Petoez, P. & Farmakalidis, E. (1995). A satiety index of common foods. *European Journal of Clinical Nutrition* **49**, 675–690.

Peryam, D. R. & Pilgrim, F. J. (1957). Hedonic scale method of measuring food preferences. *Food Technology* **11**, 9–14.

Rolls, B. J., Fedoroff, I. C., Gutherie, J. F. & Laster, L. J. (1990). Foods with different satiating effects in humans. *Appetite* **15**, 115–126.

Teghtsoonian, M., Becker, E. & Edelman (1981). A psychophysical analysis of perceived satiety: Its relation to consumatory behavior and degree of overweight. *Appetite* **2**, 217–229.