Review article

The use of visual analogue scales to assess motivation to eat in human subjects: a review of their reliability and validity with an evaluation of new hand-held computerized systems for temporal tracking of appetite ratings

R. J. Stubbs[1]*, D. A. Hughes[1], A. M. Johnstone[1], E. Rowley[1], C. Reid[2], M. Elia[3], R. Stratton[3], H. Delargy[4], N. King[4] and J. E. Blundell[4]

[1]*Rowett Research Institute, Greenburn Road, Bucksburn, Aberdeen AB21 9SB, UK*
[2]*Biomathematics and Statistics, Scotland, Rowett Research Institute, Greenburn Road, Bucksburn, Aberdeen AB21 9SB, UK*
[3]*Clinical Nutrition Group, The Dunn Clinical Nutrition Centre, Hills Road, Cambridge, UK*
[4]*Biopsychology Group, Psychology Department, University of Leeds, Leeds LS2 9JT, UK*

(Received 13 May 1998 – Revised 9 November 1999 – Accepted 22 November 1999)

This present paper reviews the reliability and validity of visual analogue scales (VAS) in terms of (1) their ability to predict feeding behaviour, (2) their sensitivity to experimental manipulations, and (3) their reproducibility. VAS correlate with, but do not reliably predict, energy intake to the extent that they could be used as a proxy of energy intake. They do predict meal initiation in subjects eating their normal diets in their normal environment. Under laboratory conditions, subjectively rated motivation to eat using VAS is sensitive to experimental manipulations and has been found to be reproducible in relation to those experimental regimens. Other work has found them not to be reproducible in relation to repeated protocols. On balance, it would appear, in as much as it is possible to quantify, that VAS exhibit a good degree of within-subject reliability and validity in that they predict with reasonable certainty, meal initiation and amount eaten, and are sensitive to experimental manipulations. This reliability and validity appears more pronounced under the controlled (but more artificial) conditions of the laboratory where the signal : noise ratio in experiments appears to be elevated relative to real life. It appears that VAS are best used in within-subject, repeated-measures designs where the effect of different treatments can be compared under similar circumstances. They are best used in conjunction with other measures (e.g. feeding behaviour, changes in plasma metabolites) rather than as proxies for these variables. New hand-held electronic appetite rating systems (EARS) have been developed to increase reliability of data capture and decrease investigator workload. Recent studies have compared these with traditional pen and paper (P&P) VAS. The EARS have been found to be sensitive to experimental manipulations and reproducible relative to P&P. However, subjects appear to exhibit a significantly more constrained use of the scale when using the EARS relative to the P&P. For this reason it is recommended that the two techniques are not used interchangeably.

Visual analogue scales: Electronic appetite rating systems: Appetite: Hunger

A specific advantage of studying the behaviour of human subjects (relative to animals) is that human subjects can be asked a number of questions relating to their motivation, sensations and attitudes. Psychologists and clinicians have long used subjective feelings of bodily sensations or functions to help in research investigations and patient management. Such assessments have been carried out in diverse conditions to examine a variety of ‘functions’: quality of life (Hunt *et al*. 1981), pain (Ohnhaus & Alder, 1975; Downie *et al*. 1978), sex, libido, depression, anxiety (Keys *et al*. 1950), nausea and appetite (Hill & Blundell, 1982). Freyd (1923) has pointed out that such ratings are the only practical equivalents of objective measurements for many types of psychological phenomena, especially introspective or verbally reported data.

Attempting to understand the role of food and other

Abbreviations: EARS, electronic appetite rating systems; P&P, pen and paper; VAS, visual analogue scale.

* **Corresponding author:** Dr R. James Stubbs, fax +44 (0) 1224 715349, email J.Stubbs@rri.sari.ac.uk

environmental influences on human appetite and feeding behaviour is now considered a crucial part of the study of energy balance in human subjects. At the present time there is considerable concern and interest in understanding how diet, drugs and behaviours such as exercise influence patterns of food and energy intake. In animal models it is only possible to study the behaviour of the subject within the constraints of the experimental design. In such studies changes in either amount or type of food eaten or both are taken as indicative of the motivation of the animal to eat. In human subjects, additional information can be obtained by studying the effects of various agents on subjectively expressed appetite, as well as feeding behaviour and energy balance. With human subjects it is possible to ask structured questions about their motivation to eat or not eat, which type of food and what amount they would consume.

The value of psychometric tools in studies of food, energy and nutrient intake

It might be supposed that the measurement of subjective motivation to eat in terms of hunger, satiety and expressed appetite for certain foods is a superfluous measurement since the critical measure is objective food intake. However, it is important to recognize that 'objective food intake' may not be as objective or as 'uncontaminated' an outcome as it ostensibly appears. Studies of human feeding often allow subjects to feed entirely *ad libitum* within the constraints of the experimental design. This can take the form of a single test meal of a constant composition, a range of food items, or subjects recording their own intakes in their natural setting. There are several other means of measuring food intake and the reader is referred to detailed reviews which describe many of the experimental limitations to the measurement of food energy and nutrient intake (Spitzer & Rodin, 1981; Hill *et al.* 1995; Stubbs *et al.* 1998*b*). The basic point is that each experimental design which measures food intake is constrained by the design characteristics of the foods or diets used. In some experiments, design constraints only allow subjects to increase or decrease the amount of a diet of constant composition that they eat (Lissner *et al.* 1987; Stubbs, 1995). Under these conditions assessments of motivation to eat can yield important information that is not made available by measuring food intake alone. For instance, in a recent study subjects were allowed to feed *ad libitum* on high carbohydrate diets of high ($<$ 6500 kJ/kg) or low ($<$ 3000 kJ/kg) energy densities. Subjects could only alter the amount they ate of a diet of constant composition for 2 weeks per treatment (Stubbs *et al.* 1998*a*). Subjects ate similar amounts of food on each diet, and energy intakes were 8·7 and 14·8 MJ/d on the low and higher energy-dense diets respectively. Measuring food intake alone suggested that simple alterations of energy density alters body weight. However, measurement of hunger showed that all subjects were far more hungry on the less energy-dense diet. This suggests that over longer periods they would have compensated by eating more or if given access to a range of foods they would have selected more energy-dense foods. Thus, gathering information on subjective motivation to eat in this study overcame some of the limitations of only measuring food intake as the outcome. This example shows how subjective and psychometric data yield important information that enables the investigator to interpret better the observed feeding behaviour.

The measurement of motivation to eat can also be of critical importance under conditions where feeding is not *ad libitum* or where voluntary energy intake is constrained, for instance, during the administration of parenteral nutrition or during therapeutic energy restriction. In situations such as this, the effect of the intervention itself on motivation to eat is of special interest. For instance, Stratton *et al.* (1998*a*) have recently described how some subjects receiving parenteral nutrition at home feel particularly hungry at meal times. This occurs despite the administration of energy and nutrients which meet their requirements. Under these conditions, the measurement of motivation to eat yields critical data that could never be inferred from measures of voluntary intake. Similarly, during weight loss therapy the measurement of energy and nutrients that have been prescribed during a period of energy restriction will tell the investigator little about the probable success of that treatment. Measurement of motivation to eat provides some of that information by informing the investigator whether subjects are able to tolerate and comply with the treatment in real life.

The measurement of motivation to eat in studies of food intake also allows different aspects of motivation to be assessed without contaminating the main behavioural outcome being measured. For instance, the effect of an experimental manipulation on intake at a test meal can be assessed. Insights into likely feeding behaviour in real life can also be obtained by assessing subjects' appetite for other foods that are not present in the experimental design.

Assessment of motivation to eat gives greater insights into feeding behaviour than can be derived from measures of voluntary intake alone. This is a critical factor if we are interested in the behaviours that alter energy balance and food choice as well as being interested in the energy and nutrients that are ingested in laboratory experiments and real-life interventions.

What are hunger, appetite and satiety?

When studying the physiology of appetite control it is important to remember that hunger, appetite and satiety have objective (unconditioned, or physiological) and subjective (conditioned, or learned) components. In order to dissociate these components it is important to attempt to measure them. Clearly this requires a means of assessing subjective motivation to eat. But what exactly are we measuring? It is important to consider what people mean when they use the terms hunger, appetite and satiety. Blundell (1979) has argued that hunger, appetite and satiety are not tangible or even objectively observed phenomena: '… instead they are explanatory concepts which are inferred from events which cannot be directly observed or defined …' by people at large. In other words, people use these terms as an aggregate description of several sensations they recognize as predictors of their normal behaviour. The actual sensation of hunger is therefore likely to differ both quantitatively and qualitatively between individuals, as shown in a study carried out by Monello & Mayer (1967). A questionnaire completed by 603 men, women, boys and girls, which investigated the individuals' subjective experience of hunger and satiety, could find no clear constellation of traits, sensations or characteristics that typified hunger in most subjects (Monello & Mayer, 1967). Different people feel hungry in different ways. The amount of perceived hunger or appetite that triggers feeding will differ between individuals in a given situation and will differ within an individual in different situations. This suggests that these motivations are best measured in within-subject, repeated-measures designs. Blundell (1979) thus argues that hunger, appetite and satiety are hypothetical constructs that we use to conceptualize our perception of sensations or motivations which are themselves indirectly linked to processes which influence our behaviour. In other words, physiological signals, environmental influences and past experience summate to determine the multiple afferent inputs which lead us to 'feel' like eating or not eating. However, the source of these inputs (e.g. low blood glucose, a social occasion, or previous food poisoning respectively) are not often directly measurable as indicators of a need for food or a need to avoid food. We therefore invoke words as a common reference which conceptualize our inclination to behave in given ways, but do not actually describe every detailed sensation underlying that expressed motivation. For instance, we do not say 'I feel tired, empty, unable to concentrate, irritable and impatient to eat', we say 'I'm hungry'. As hunger is a subjectively expressed construct that is used to express a motivation to eat, the most appropriate measure of hunger is its subjective expression at a given time. The same is true for other aspects of motivation to eat. Tracking changes in subjective hunger (or other aspects of appetite) over time provides important information in relation to the structure or the effects of feeding events, for instance, the effects of diet composition on feeding behaviour or the effects of physiological variables on the appetite control system. How are hunger and appetite measured?

The use of visual analogue scales in appetite research

A number of systems has been devised to ask subjects specific questions relating to aspects of their motivation to eat (Hill & Blundell, 1982; Silverstone & Goodall, 1986; Leathwood & Pollet, 1988; De Graaf, 1993). One of the most productive and popular systems is the use of VAS. The VAS have become particularly popular in pain (Ohnhaus & Alder, 1975; Downie *et al*. 1978) and appetite research (Hill & Blundell, 1982; Leathwood & Pollet, 1988; De Graaf, 1993). The VAS typically take the form of a straight line with two extreme states anchored at either end. For example, a scale with a question such as 'how hungry are you?' would be anchored with 'not at all hungry' at one end and 'as hungry as I have ever felt' at the other end. Such scales have the advantage of being easy and quick to use, simple to interpret, they do not require the subject to invoke their own descriptive terms, they allow considerable discrimination and they are presented in a standardized format that can be compared under a variety of different experimental manipulations.

The most commonly used methodology in the UK is that of Hill & Blundell (1982). The original version of this questionnaire contained six questions: 'How hungry do you feel?' (not at all hungry/as hungry as I've ever felt); 'How full do you feel?' (not at all full/as full as I have ever felt); 'How strong is your desire to eat?' (very weak/very strong); 'How much do you think you could eat now?' (nothing at all/a large amount); 'Urge to eat' (no urge to eat/strong, want to eat now, waiting is very uncomfortable); 'Preoccupation with thoughts of food' (no thoughts of food/very preoccupied difficult to concentrate on other things). These ratings have been found to co-vary to a great extent and it is therefore pertinent to ask whether the six questions used relate to a single phenomenon, motivation to eat, or to more than one underlying motivation or process. In several studies, with the numbers of subjects ranging from five to sixteen, we have applied principal components analysis to the six commonly used 100 mm VAS to identify distinct dimensions in the responses to the questions (Reid *et al*. 1998). In almost every case the first principal component is essentially an average of the six VAR. In every study this component explains at least 85 % of the variation observed across the six VAR and can be thought of as a general measure of satiety or appetite. Unfullness is the maximum measurable value of the scale (100) minus the recorded value for fullness. Subjects usually rate fullness as the opposite of hunger and the other four questions; thus in principal components analysis, it is most useful to describe this similar (but reverse) pattern as 100 (maximum value) minus fullness (100 – fullness). The analysis reveals a pattern similar to that for other ratings. Principal components analysis also suggests a second compound rating, contrasting a gut-based sensation, as measured by unfullness, with motivation to eat, as measured by desire, urge, prospective consumption and thoughts of food. In each study, the first two principal components explain over 90 % of the variation. Thus, if these data are generally applicable to large numbers of subjects, it would appear that questionnaires used to monitor hunger could be redesigned to account for these two aspects of hunger and to assess other possible dimensions of motivation to eat and drink. These analyses suggest that these six VAS collectively relate to a general motivation to eat and a sense of gastrointestinal repletion which subjects appear to distinguish from their general motivation to eat. Future questionnaires could amplify these distinctions and add other aspects of motivation to eat not included in these particular assessments. Having considered why we should measure hunger and satiety, what they are and what the VAS most commonly used in the UK appear to identify in motivation to eat, it is pertinent to consider the reliability and validity of VAS.

The reliability and validity of visual analogue scales

A major problem with psychometric ratings of sensation or motivation is that it is difficult if not impossible to objectively validate them. Freyd (1923) noted that in choosing which sort of rating to use it would be ideal to compare different ratings with each other in relation to an objective measure of the phenomenon under scrutiny. However, as

Freyd himself recognized, if this were the case, a rating would not be required in the first place. Freyd invoked two basic categories of criteria to assess the value of different ratings scales, (1) their practical validity: ease of use, interpretability, convenience and universality of the scale and (2) statistical validity: statistical comparison with other ratings systems for a given trait, test–retest validity (repeatability), the normality and spread of data obtained from ratings. Too low a spread of data means that the rating has little discriminatory power, too high a degree of variability and the rating may be imprecise or 'noisy' in the assessments made when using it.

As stated above the reliability and validity of a rating scale can be difficult to demonstrate unequivocally since there are often few if any objective measures of the trait with which to compare the rating system. However, there are comparisons which can be made to assess the reliability of an appetite rating system: (1) the apparent validity of the rating in terms of its ability to predict the behaviour which is being assessed by the rating is one method of demonstrating its usefulness; (2) the rating can be compared under conditions where it should change if sensitive (e.g. nutrient loads, anorectic drugs); (3) the reproducibility (test–retest–reliability) of the rating can also be assessed.

Predictive power

de Castro & Elmore (1988) have reported significant correlations between subjectively-rated hunger and the reported energy content of eating bouts, in nine men and twenty-two women self-recording their food intakes for 7 consecutive d. Subjects also recorded their subjective hunger at the beginning of each meal on a 7-point scale. Pearson product moment correlations were calculated between reported hunger, energy intake at the meal (defined as intakes >0·21 MJ) and the interval since the last meal. The authors found that the amount eaten in a meal was correlated positively with subjective hunger (r 0·27; $P < 0·05$), as was the duration of the interval preceding the meal (r 0·14; $P < 0·05$). de Castro & Elmore (1988) note that while consistent and significant, these average within-subject correlation coefficients were small. They caution however that these relationships were established under free-living conditions where a number of uncontrolled and unmeasured variables constitute noise which inflate error variance estimates. Under more controlled conditions these correlation coefficients should (theoretically at least) be higher.

Mattes (1990) has examined the relationship between subjective hunger (tracked hourly during waking hours) and feeding behaviour under free-living (real life) conditions in twelve men and twelve women self-recording food intake and subjective hunger, over 7 consecutive d. When hunger ratings at the start of each hour were correlated with reported intake in the hour following each hunger rating, the group correlation between hunger ratings and energy intake was r 0·50 ($P < 0·02$). This effect was apparent for the whole group of subjects on five weekdays but not the two weekend days. Thus for the group subjective hunger was a reasonably good predictor of when eating will occur under these conditions where many extraneous factors may influence or obscure the relationship between hunger and feeding. When these analyses were conducted within individuals there were no significant associations between hunger ratings and energy content of eating occurrences in the same hour. Eating occurrence in this study was defined as a reported intake >0 MJ. The within-subject effect of eating on hunger was also assessed. Hunger declined in the time points before and after an eating episode on 49 % of eating occasions. When subjects fasted during the same time interval, Mattes (1990) reports that hunger increased 46 % of the time. However, hunger ratings were not a useful proxy measure of either computed energy content of eating occurrences or total reported number of eating occurrences (Mattes, 1990), indicating that they cannot be used as a proxy for quantitative variables such as energy intake. This is not surprising since in the natural setting people often eat when they are not hungry and sometimes do not eat when they do feel hungry. It is important to note that the studies of de Castro & Elmore (1988) and of Mattes (1990) took reported energy intakes as given. These are not reported in the documents and so it is difficult to assess their plausibility in relation to expected energy requirements. Only recently has the likely impact of mis-reporting of dietary intakes been fully appreciated. Mis-reporting is likely to contribute further to the inflation of error variance estimates mentioned by de Castro & Elmore (1988). It therefore appears that subjectively-rated hunger is a valid predictor of human feeding behaviour in terms of approximate meal size and frequency under free-living conditions, but a poor proxy of quantitative variables such as energy intake. Mattes (1990) notes that when fasted or given pharmacological treatments which affect appetite, subjects display changes in motivation to eat that predict changes in behaviour consistent with the effects of the intervention (Speigel *et al*. 1987; Blundell & Hill, 1988; Wolkowitz *et al*. 1988). However, there is a number of other less extreme experimental conditions in which hunger does not reliably predict feeding behaviour (Pi-Sunyer *et al*. 1982; Thompson *et al*. 1982; Trenchard & Silverstone, 1983; Rogers *et al*. 1988; Rolls *et al*. 1988). Mattes (1990) argues that these discrepant findings can be resolved by the argument that hunger is an innate reflection of physiological systems controlling appetite (which becomes apparent when those systems are experimentally manipulated) and also an entrainable response under more normal dietary conditions. In some experiments changes in hunger may reflect previous entrained responses more than they indicate a response to (often modest) nutritional challenges which comprise the novel experimental intervention.

De Graaf (1993) has cautioned that the correlation coefficient between a subjective expression of appetite and energy intake is heavily influenced by the exact statistical relationship that is calculated. For instance, hunger gradually climbs as the time since the last meal increases. As the onset of the next meal approaches there is often a more rapid elevation of hunger. Therefore correlations between hunger in the hour preceding a meal and amount eaten at the meal will be greater than those calculated between average hourly hunger in the intermeal interval and energy intake at the next meal. Different authors tend to calculate different statistical relationships. For example, the definition of a meal given by Mattes (1990) differed from that of de Castro & Elmore (1988). Additionally, assessments have tended to examine the relationship between appetite and energy (but not food) intake (i.e. amount eaten).

The variation in the subjectively-expressed hunger of human subjects (using hourly ratings) is heavily influenced by time which often accounts for 20–30 % of the variance in our experiments. This suggests that hunger itself exhibits a large learned component in relation to entrained feeding patterns (Mattes, 1990; De Graaf, 1993; Blundell & Stubbs, 1997). Furthermore, the factor which tends to account for most of the variance in VAS of a group of subjects is inter-subject variation in response, which has been found to account for >50 % of the variance in some of our experiments. This means that VAS are best used in within-subject comparisons. Where small groups of subjects are being studied, and a significant treatment effect found, it is useful to consider what proportion of subjects were accountable for the significant group effect. In addition to assessments of predictive power, the reliability and validity of VAS can be evaluated in terms of their sensitivity to experimental manipulations.

Sensitivity to experimental manipulations

VAS in appetite research have been repeatedly found to be sensitive to a number of experimental manipulations including alterations in diet composition (de Castro & Elmore, 1988; Hill & Blundell, 1992; Johnstone *et al.* 1996; Stubbs *et al.* 1996), alterations in energy intake (Westrate, 1992; De Graaf *et al.* 1993; Stubbs *et al.* 1997) and administration of drugs which stimulate (Thompson & Campbell, 1977, 1978) or inhibit appetite (Hill & Blundell, 1990). Furthermore, changes in expressed appetite using VAS have been related to changes in physiological variables thought to exert important influences on appetite. For example, the postprandial utilization of carbohydrate may influence hunger (Mayer, 1955; Campfield *et al.* 1992; Raben *et al.* 1995). It has been noted that a number of relatively modest experimental manipulations have not shown subjective hunger to predict feeding behaviour, possibly because the experiments do not influence innate physiological systems to a great extent. We have found in our laboratory that VAS are often more sensitive to a dietary manipulation than are changes in food and energy intake (e.g. Johnstone *et al.* 1996; Stubbs *et al.* 1998*a*), at least when subjects are given *ad libitum* access to an unfamiliar diet of fixed composition, so that they can alter the amount but not the composition of foods they eat. This can be taken to indicate sensitivity since a degree of motivation usually accrues before a behavioural response is initiated. As mentioned above it appears that the correlation between manipulations of energy balance and hunger are stronger when those manipulations exceed or disrupt the entrained relationship between hunger and feeding that characterizes subjects' usual patterns of behaviour.

Test–retest reliability

It is difficult to examine the reliability (i.e. reproducibility) of VAS in terms of between-subject reliability as lack of correspondence between scores may be due to genuine differences in individuals' interpretation of the scale. It is possible to assess the test–retest–reliability of VAS in within subject comparison. We have found that when subjects are fed to energy balance on fixed meals at fixed meal times over consecutive days or non-consecutive days, their ratings of hunger show a consistent 'peak and trough' pattern which is highly reproducible within subjects (see later). However, caution is necessary in assessing the test–retest reliability of ratings since if subjects were fed in energy imbalance to the same extent over consecutive days or if imbalances occurred on intervening days, one would not expect a good test–retest relationship. For instance, twelve male subjects participated in a recent study which followed a 3 d sequence. On day 1 (maintenance period) subjects were given a fixed intake of a medium-fat diet (40 % fat, 47 % carbohydrate and 13 % protein as a proportion of dietary energy) fed to energy requirements (1·6 × resting metabolic rate). On day 2, they underwent a total fast. On day 3 they had *ad libitum* access to a selection of high-protein, high-carbohydrate or high-fat foods. Average daily VAS scores were 32, 61, 42 (SED 4) mm on days 1, 2 and 3 respectively. Energy intakes were similar on days 1 and 3 but hunger was significantly different due to the intervening energy imbalance.

Leathwood & Pollet (1988) examined the effects of slow-release starches on plasma glucose and hunger in six subjects who were given six different test meals in triplicate. On each occasion hunger sensations were rated on a 10-point scale. The pooled within-subject, within-treatment standard deviations for satiety, fullness, hunger and 'gourmandize' (a term used by Leathwood & Pollet (1988)) were 2·09, 2·06 and 2·1 and 2·1 respectively, showing that within a given treatment, ratings were consistent. Within-subject variability in ratings increased over time (Leathwood & Pollet, 1988). Lappalainen *et al.* (1993) also found no significant difference between VAS ratings measuring appetite after three identical meals on three separate occasions. Raben *et al.* (1995) note that this in itself does not strictly constitute a test of reproducibility, citing Bland & Altman's (1986) test for assessing agreement between two methods in clinical measurements. Raben *et al.* (1995) found that subjective appetite scores are not easily reproduced when tested after identical (albeit highly unusual) meals in the same subjects on different days under standardized conditions (Raben *et al.* 1995).

Silverstone & Goodall (1986) looked at the between-group reproducibility of VAS scores using an anorectic drug (Triflorex) and found it to exert similar effects in suppressing hunger in two independent study populations. Any such comparisons would, however, be of a semi-quantitative nature due to a large inter-subject variability in use of the VAS. Similarities or differences between the average scores of different groups are heavily affected by inter-subject variability in response when using the VAS. Hence comparisons are semi-quantitative because like is not being directly compared with like.

We have recently demonstrated the test–retest reliability of VAS using paper and pen and new hand-held computer systems used to assess motivation to eat in a series of three experiments (Delargy *et al.* 1996; Stubbs *et al.* 1997; Stratton *et al.* 1998*a*). While there were some differences between techniques the VAS were largely reproducible both within and between techniques.

The previous section reviews evidence that subjective motivation to eat predicts feeding behaviour in free-living subjects going about their normal routines in their familiar environmental setting. There is strong evidence that a large component of motivation to eat is entrained and so contrary to expectation it is not always easy to demonstrate a clear relationship between hunger and manipulations of energy balance in the laboratory. This is particularly so where small manipulations of energy balance are used. They are simply not large enough to overcome entrained effects. Given these *caveats*, there is evidence that subjective motivation to eat is sensitive to experimental manipulations. There is also evidence that subjective ratings of motivation to eat are reproducible. With the development of new techniques to improve data capture, using VAS, further studies have been conducted which provide evidence of the sensitivity and reproducibility of VAS used to assess motivation to eat.

Recent developments in data collection using visual analogue scales

There are problems inherent in the use of the traditional pen and paper (P&P) method of VAS. In many experiments subjects will be largely unsupervised while completing questionnaires and the integrity of the data cannot be verified. We have had to discard data on two occasions when individuals have completed all their questionnaires at the end of the day, rather than hourly throughout the day. It is not surprising that subjects often forget to complete a questionnaire on the hour, and studies over several days are often characterized by data-sets that contain numerous missing values. Under these conditions statistical tests must be conducted to check if the number and distribution of missing data values are balanced across time and treatments. Interestingly, studies conducted over similar time periods in the controlled environment of the calorimeter yield more complete data-sets (Stubbs *et al.* 1995). Other errors may arise in the completion of the VAS. Subjects can fail to record the time or other important details. Tabulation and analysis of the data is very time consuming when using paper and pen VAS since each question has to be physically measured and recorded by the investigator. Errors may arise when transposing the data from the paper onto a spreadsheet. With these considerations in mind some groups have now developed automated systems for data collection. Electronic appetite rating systems (EARS) have recently been developed to overcome the limitations and constraints of the P&P method. Personal computer-based electronic appetite ratings systems have been developed by Yeomans *et al.* (1997) using Macintosh desktop computers. In addition to desktop systems two portable systems have recently been developed. One system employs the Psion organizer (Psion plc, London, UK) (Delargy *et al.* 1996) and the other employs the Apple Newton Message Pad (Apple Computer Inc., Cupertino, CA, USA) (Stubbs *et al.* 1997; Stratton *et al.* 1998*b*). The Psion and Newton were specifically chosen as electronic devices that mimic the ease of use of P&P, as the Psion consists of a moving cursor along a dotted line and the Newton consists of a pen-based graphical interface which is ideal for use with VAS. A program has been written for the Psion which produces a VAS on the graphical display (© J. E. Blundell and N. King). An object-based program has been written for the Newton (© R. J. Stubbs and M. Elia). The program makes use of the Newton's features to prompt the subject hourly to complete a series of VAS which appear on the screen in sequence. Both machines have been developed with the following features: entries are automatically date and time verified; incomplete and/or incorrect entries and back reference to previous entries are precluded. The Psion and Newton store the data in a format which can be readily downloaded and the questions asked can be changed. These devices have recently been described and validated as machines that can greatly enhance the ease and quality of time-verified data capture (Delargy *et al.* 1996; Stratton *et al.* 1997; Stubbs *et al.* 1997). One study has now been conducted on the Psion (Delargy *et al.* 1996) and two independent studies have been conducted on the Newton-based EARS (Stubbs *et al.* 1997; Stratton *et al.* 1998*b*), in comparison with the traditional P&P method. In addition other studies have now used the Newton-based EARS in dietary manipulations which have demonstrated its sensitivity (e.g. Stubbs *et al.* 1998*c*). Some of the comparisons of the EARS with traditional P&P involved comparison of reproducibility and validity, and these studies provide further evidence of the suitability of using VAS in appetite research. In all of these comparisons a series of questions was put to subjects hourly, relating to motivation to eat. The methodology used VAS and questions as or similar to those described by Hill & Blundell (1982).

Delargy *et al.* (1996) conducted a study which compared the responsiveness of the Psion-based EARS or P&P with three lunches which were low energy (1·6 MJ), high energy (3·2 MJ) or high energy–high fat (3·2 MJ). This study was conducted in sixteen men and women, as a 3×2 design (three lunch conditions and two measurement techniques, Psion EARS and P&P). Each subject was thus studied on six separate occasions in a counterbalanced order. Ratings were recorded hourly during waking hours. In this experiment subjects consumed a standard breakfast (at 08.30 hours), the manipulated lunch (at 12.30 hours) and eating was prohibited until 17.30 hours, after which subjects recorded in food diaries everything they ate. This experiment compared subjects' response with: (1) differing amounts of food and energy and (2) differing energy densities, of lunches containing the same energy content.

Delargy *et al.* (1996) found that the mean ratings for the two techniques were very similar, although the standard deviation about the mean was slightly lower with the Psion-based EARS than P&P. Around 50 % of the data for the EARS fell within a slightly narrower range than for the P&P. Within-subject linear regression of EARS (outcome) against P&P gave r^2 values for various aspects of motivation to eat ranging from 65 to 76 %. Parenthetically, the two methods gave far lower r^2 values for questions relating to lethargy (r^2 22 %). ANOVA of the ratings before the manipulated lunch also showed there to be no effect of method of measurement or sex of subject on ratings of motivation to eat. These results indicated that the methods were comparable before the nutritional challenge. ANOVA and *post hoc* tests revealed that during the period after lunch before *ad libitum* intake, significant differences between the lunches occurred, depending on whether EARS or P&P was used. Both methods detected significant differences between the low- and high-energy lunches. The EARS detected significant differences between the high-energy and high-energy–high-fat lunches only in the first hour after lunch, whereas the P&P did not. These effects were however, subtle, since ANOVA revealed no overall effect of method on the response to the differing lunches. The results showed that both of the techniques were sensitive to nutritional manipulations that typify single-meal manipulations made in the laboratory and were of a similar magnitude to changes in feeding in real life. The two techniques were largely reproducible with the notable exception that subjects tended to exhibit a slight but significantly more constrained variance when using the EARS. This effect was clearly not sufficient to affect differently the recorded response of subjects to nutritional challenges in the laboratory.

Stratton *et al*. (1998*b*) have compared the response of subjects using the Newton-based EARS with the P&P. The study was conducted in free-living subjects consuming their normal diet *ad libitum*. The purpose of the study was to examine the results of each method when rated at the same time (in consecutive order) in twelve subjects and to evaluate the test–retest reliability of Newton-based EARS *v*. P&P in eight subjects. In the direct comparison seven women and five men completed the EARS and the P&P consecutively, every waking hour, for 2 consecutive d. To assess the test–retest reliability of the two methods eight subjects continued for a further 2 d. In random order on one day they completed two copies of the P&P questionnaire, and on the other day two copies of the EARS questionnaire, in the same manner as described for the EARS *v*. P&P comparison.

We used the method of Bland & Altman (1986) for the comparison of two measurement techniques to assess whether there was bias between the two means of measuring subjective hunger (Stratton *et al*. 1998*b*). In the P&P *v*. EARS comparison there was no significant difference between the two techniques for responses of motivation to eat, except for hunger and fullness ratings. For hunger and fullness the differences between the two techniques became significant at the extreme ends of the VAS (i.e. very high and very low ratings). Indeed for all of the responses there was a tendency for the differences between methods to become larger with increasing score, although the overall pattern of change and sensitivity of the two methods remained similar. This effect was greatest (and hence significant) for the hunger and fullness ratings. The difference in the use of the two techniques appeared to be due to a slightly more constrained use of the range of responses when using the EARS, giving a correspondingly constrained variance in response for this technique. This means that the bias was apparent in the following directions. The highest scores on the EARS were lower than the highest scores on the P&P and the lowest scores on the EARS were higher than the lowest scores on the P&P, i.e. the variance was constrained with reference to the centre of the VAS. In the test–retest comparison for both techniques there was no significant difference in the individual hourly results or in the mean summary appetite ratings for the whole study period. There was significant bias for one response: ‘How strong is your desire to eat?’ on the EARS. Nevertheless the pattern of change remained similar for this response. This evaluation of the Newton-based EARS *v*. the P&P suggested that both techniques are largely reproducible and follow the same pattern of response. However, as for the comparison of the Psion-based EARS with P&P (Delargy *et al*. 1996), the electronic version tended to be used by subjects with a slightly and significantly more constrained range of responses. It is this tendency for subjects to avoid the very extreme ends of the shorter electronic VAS, which contributes the greater disparity between methods at the extremes of the response range (e.g. ‘not at all hungry’ or ‘as hungry as I have ever been’).

Stubbs *et al*. (1997) have recently compared the response of subjects using the Newton-based EARS and the P&P, when consuming fixed mandatory meals of the same energy density, energy and nutrient content, at fixed time points over a day (breakfast 08.30 hours, lunch 13.30 hours, supper 17.20 hours). This experiment was conducted in ten men and ten women on two separate days. On each day the diet

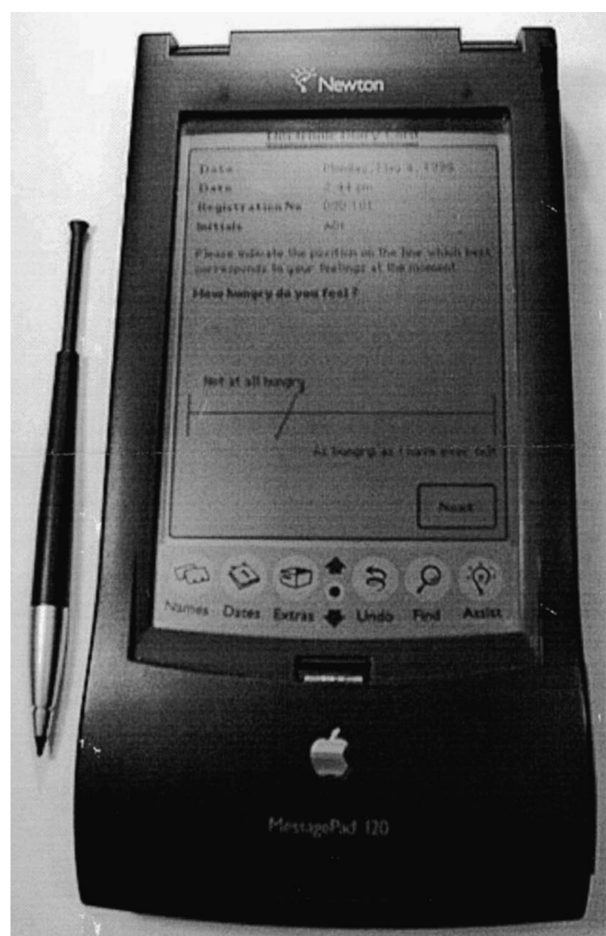

Fig. 1. The Newton palmtop computer showing visual analogue scales.

was designed to match energy requirements at $1{\cdot}5 \times$ BMR. The twenty subjects were each studied once in a randomized study design which lasted 4 d. Day 1 was followed by two intervening days (days 2 and 3; 'intervening period') and then another test day (day 4). Both the electronic and P&P methods were completed hourly but the order in which the two methods were used was alternated on even and odd hours, such that on even hours subjects were required to complete P&P questionnaires first and on odd hours the EARS first, every waking hour between 08.00 hours and 23.00 hours. Each subject kept to the same order on the two test days. This study examined the subjective responses to fixed intakes of energy and food at fixed times, using two different methods of measurement. In addition the test–retest reliability of each technique was compared on two non-consecutive days.

As in the Stratton study (Stratton *et al.* 1998*b*), the mean daily ratings for hunger and fullness were compared using Bland & Altman's (1986) test for bias between two measurement techniques. The mean difference between the two techniques was significantly different from zero (hunger t19 –2·26; $P = 0{\cdot}036$; fullness t19 –4·84; $P < 0{\cdot}001$). Further scrutiny revealed that this bias was due to a significantly higher variance in the ratings using P&P than the EARS (hunger t19 –6·32; $P = 0{\cdot}001$; fullness t19 –4·83; $P < 0{\cdot}001$). These data patterns were similar for the other responses but less pronounced for 'desire to eat' and 'urge to eat'.

To further explore the relationship between the expressed variables using the two techniques, linear regression analysis was conducted using the P&P technique as the predictor variable and the EARS technique as the outcome variable. Models were checked for gender, order and day effects. The results demonstrated a strong linear relationship between the two methods (r^2 76·6–98·6). The plots of mean hourly hunger for all subjects across days 1 and 4, using the paper and EARS are given in Fig. 1. Regression analysis showed that for all responses, the slope of the line was less than 1 and the constant was significantly greater than zero ($P < 0{\cdot}01$). This indicates that a smaller range of values tended to be used for the EARS than the P&P, both being 100 unit scales of a different absolute size (66 mm *v.* 100 mm respectively). As in the studies of Delargy *et al.* (1996) and Stratton *et al.* (1998*b*), subjects tended to avoid the extremes of the scale on the EARS and thus the highest scores for the EARS tended to be lower than the highest scores for the P&P and the lowest scores for the EARS tended to be higher than the lowest scores for the P&P.

Fig. 2 illustrates the reproducibility in patterns of hunger using the Newton on days 1 and 2. The lines with symbols are mean hunger ratings at each time point smoothed using a smoothing spline with 4 degrees of freedom. The lines around the margins are reference bands. At each time point the region has a width of twice the standard error of the differences in hunger between the two days, and is centred at the average of the two means. Where the plotted means exceed this band, the means are significantly different, since the means exceed 2 SE of the difference. All the mean ratings for all hours of the day lie within the reference bands, indicating that hunger ratings were not significantly different at any time during days 1 and 2 of the experiment.

In addition, the Newton-based EARS has been used to

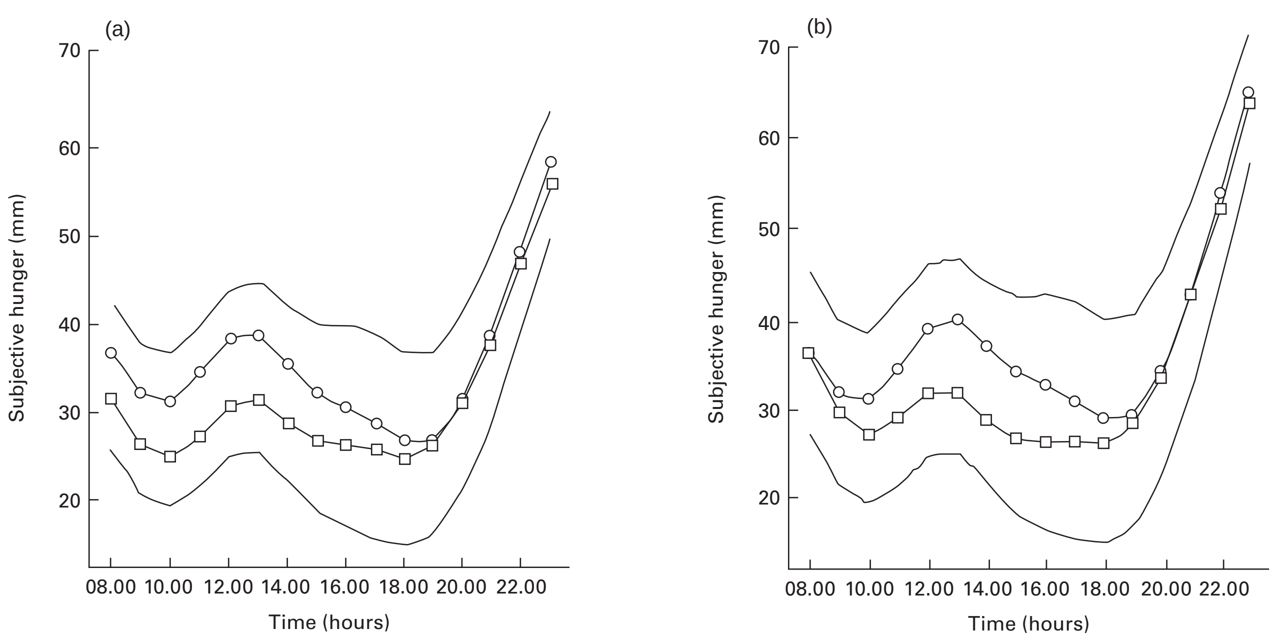

Fig. 2. Reproducibility of subjective hunger ratings between day 1 and the second day (day 4) of the experiment in twenty subjects using (a) the Newton palmtop computer or (b) paper and pen. (□—□, ○—○), mean hunger ratings on day 1 and day 4 respectively, smoothed using a smoothing spline with 4 degrees of freedom. The margins are reference bands corresponding to twice the standard error of the differences in hunger between the two days, and are centred at the average of each pair of mean values. If plotted means exceed this band they are significantly different, since the means exceeded 2 SE of the difference. All values lie within the reference bands indicating the reproducibility of the ratings under the same conditions on non-consecutive days.

test for differences in subject response to nutritional challenges. While these studies were not used as a comparison to P&P, they did use relatively subtle nutritional manipulations and significant differences were detected between those experimental conditions when the subject used the EARS. In one of these studies (Stubbs *et al.* 1998*c*), sixteen men were each studied four times in a randomized counterbalanced design. On each occasion they received a mandatory intake as breakfast (08.30 hours) plus a snack at 10.30 hours. This comprised 80 % of resting energy requirements as high-protein, high-carbohydrate, or high-fat foods (60 % of energy in each case) or an equal mixture of macronutrients (33 % of energy from protein, from carbohydrate and from fat). All treatments contained the same energy content and had the same energy density and thus only differed in macronutrient content. From 12.30 hours onwards subjects had *ad libitum* access to the selection of thirty high-protein, high-carbohydrate and high-fat foods. Subjects were significantly less hungry before lunch on the high-protein and mixed (33 % protein) treatments ($F_{3,44}$ 7·35, $P < 0{\cdot}001$). They were also significantly more full on the high-protein treatment and significantly less full on the high-fat treatment relative to all other treatments. At lunch subjects ate significantly more on the high-fat treatment than any other treatment. This study has shown that when all other factors (including energy density) are held constant and only the macronutrient content of foods is manipulated, the EARS is sensitive to subtle nutritional challenges. Furthermore, motivation to eat predicted feeding behaviour in a manner consistent with the effects of the manipulated breakfasts. These studies collectively suggest that the EARS has considerable utility for increasing reliability of data collection and reducing investigator workload. How do subjects themselves feel about this new technique?

Subject attitudes to the electronic appetite rating systems and paper and pen techniques

It is important to ascertain the attitude of subjects to the two techniques. While the use of the EARS may have lightened the load of the investigator it will not improve the quality or reliability of data collection if subjects are less able or willing to use it. The three comparative studies have also evaluated the preference of subjects for the EARS over the P&P technique (Delargy *et al.* 1996; Stubbs *et al.* 1997; Stratton *et al.* 1998*b*). EARS-based questionnaires have previously been developed for use in self-rating of symptoms by patients in the clinical setting. The questionnaires used were also evaluated against traditional P&P methods in patients suffering from gastrointestinal disorders. Most of the patients (57 %) preferred the electronic method with 13 % preferring P&P and 30 % expressing no preference. Results obtained using either technique were within 2 % of each other (Drummond *et al.* 1995). In the study of Delargy *et al.* (1996) all of the subjects found the EARS easy to use, 81 % preferred it to the P&P and 75 % found it more convenient. The subjects in this study were young members of the University population, whereas in the studies of Stratton *et al.* (1997) and Stubbs *et al.* (1997) a wider range of subject types was used. In the study by Stratton *et al.* (1997) 54 % of subjects preferred the P&P, 38 % the electronic method and 8 % had no preference. In the study by Stubbs *et al.* (1997) there were no apparent differences in the subjects' (males and females combined) preference for either technique. The main difference between techniques was that women found the computer technique to be more tedious than did men. With this exception it appears that subjects had largely similar attitudes to both techniques (Delargy *et al.* 1996; Stubbs *et al.* 1997; Stratton *et al.* 1998*b*).

In summary, these studies have found that the EARS do produce similar (largely reproducible) responses relative to P&P. The main difference between subjects' use of the EARS and P&P was that they tended to avoid the very extremes of the scale when using EARS as compared with P&P. The greatest discrepancies between the two techniques would have occurred when subjects were scoring very high or low values. This effect appears to be consistent across the three studies which have compared the two techniques. Because of these differences in use of the two techniques it should be recommended that in order to maximize test–retest reliability of each technique, the EARS and P&P should not be used interchangeably. While the EARS greatly improves the ease and reliability of data capture for the experimenter, there was some variability between types of study populations in their preference for either technique. It is probable that younger or more computer-literate subjects will find the EARS easier to use than subjects less familiar with this type of tool.

While the EARS is a potentially useful tool in appetite research, there are also disadvantages to using EARS for data collection. These instruments are expensive and easily lost or stolen. Furthermore, if a subject damages or loses a unit during the course of a study all of the data contained in it will be lost also. When using P&P subjects tend to take only a small number of questionnaires with them at a time, and so tend to avoid total data loss. Therefore the data have to be downloaded daily when using the EARS.

A particular attraction of palmtop computers for data collection in studies of human feeding behaviour is that other applications can be used or developed for simultaneous or alternate data capture. For instance, a diary facility enables subjects to use the Newton as a portable food diary in addition to a psychometric data collector. Furthermore the initial program was written so that the questions asked can be changed in content and number to suit the study. The advantages of the new technique in terms of quality, reliability and speed of data collection is the adaptability and versatility of the tool. These features make it an ideal data capture tool for use in free-living subjects, the laboratory and clinical setting.

Summary and conclusion

VAS of subjective motivation to eat are psychometric tools. The results obtained from VAS are neither objective nor strictly quantitative and yield the most valuable information when combined with other aspects of feeding behaviour and energy balance. It is important to recognize that subjectively rated motivation to eat is not an inevitable outcome of underlying physiological processes. Rather it is the subject's own interpretation of their own sensations and motivations, which are influenced *inter alia* by underlying physiological processes (Blundell, 1979). A direct and high correlation between feeding behaviour and subjectively expressed hunger or appetite should not therefore always be expected. Bearing these *caveats* in mind this present paper reviews the evidence relating to the validity and reliability of VAS and finds that: (1) VAS show some ability to predict aspects of feeding behaviour and act as a useful adjunct to measures of food, energy and nutrient intake; (2) they show sensitivity to experimental manipulations, provided those manipulations exceed or disrupt the effects of conditioned motivation to eat (e.g. hunger at meal-times); (3) the VAS show good reproducibility (test–retest reliability) under controlled conditions provided that they are used in within-subject designs and provided that different VAS systems (e.g. P&P and EARS) are not used interchangeably.

Acknowledgements

This work was supported by the Scottish Office Agriculture, Environment and Fisheries Department and BBSRC grant no. F02501.

References

Bland JM & Altman DG (1986) Statistical methods for assessing agreements between two methods of clinical measurement. *Lancet* **1**, 307–310.

Blundell JE (1979) Hunger, appetite and satiety – constructs in search of identities. In *Nutrition and Lifestyles*, pp. 21–42 [M Turner, editor]. London: Applied Science Publishers.

Blundell JE & Hill AJ (1988) On the mechanism of action of dexfenfluramine: effect on allesthesia and appetite motivation in lean and obese subjects. *Clinical Neuropharmacology* **11**, S121–S134.

Blundell JE & Stubbs RJ (1998) Diet composition and the control of food intake in humans. In *International Handbook of Obesity* [GA Bray, C Bouchard and WPT James, editors]. New York, NY: Dekker Inc.

Campfield LA, Smith FJ, Rosenbaum M & Geary N (1992) Human hunger monitoring of motivation to eat. *International Journal of Obesity* **20**, Suppl. 4, 104 Abstr.

de Castro J & Elmore DK (1988) Subjective hunger relationships with meal patterns in the spontaneous feeding behaviour of humans, evidence for a casual connection. *Physiology and Behaviour* **43**, 159–165.

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

De Graaf C, Schreurs A & Blauw GH (1993) Short term effects of different amounts of sweet and non sweet carbohydrates on satiety and energy intake. *Physiology and Behaviour* **54**, 833–843.

Delargy HJ, Lawton CL, Smith FC, King NA & Blundell JE (1996) Electronic appetite rating system (EARS): validation of continuous automated monitoring of motivation to eat. *International Journal of Obesity* **20**, Suppl. 4, 104 Abstr.

Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA & Anderson JA (1978) Studies with pain rating scales. *Annals of Rheumatic Disease* **37**, 378–281.

Drummond HE, Ghosh S, Ferguson A, Brackenridge D & Tiplady B (1995) Electronic quality of life questionnaires: a comparison of pen-based electronic questionnaires with conventional paper in a gastrointestinal study. *Quality of Life Research* **4**, 21–26.

Freyd MT (1923) The graphic rating scale. *Journal of Educational Psychology* **14**, 83–102.

Hill AJ & Blundell JE (1982) Nutrients and behaviour: research strategies for the investigation of taste characteristics, food preferences, hunger sensations and eating patterns in man. *Journal of Psychological Research* **17**, 203–212.

Hill AJ & Blundell JE (1990) Sensitivity of the appetite control system in obese subjects to nutritional and serotoninergic challenges. *International Journal of Obesity* **14**, 219–233.

Hill AJ, Rogers PJ & Blundell JE (1995) Techniques for the experimental measurement of human eating behaviour and food intake: a practical guide. *International Journal of Obesity* **19**, 361–375.

Hunt SM, McKenna SP, McEwen J, Williams J & Papp E (1981) The Nottingham Health Profile: subjective health status and medical consultations. *Social Science and Medicine* **15**A, 221–229.

Johnstone AM, Stubbs RJ & Harbron CG (1996) Effect of overfeeding macronutrients on day-to-day food intake in man. *European Journal of Clinical Nutrition* **50**, 418–430.

Keys A, Brozek J, Henschel A, Mickelsen O & Taylor HL (1950) The biology of human starvation. Minneapolis, MN: University of Minnesota Press.

Lappalainen R, Mennen L, van Weert L & Mykkanen H (1993) Drinking water with a meal: a simple method of coping with feelings of hunger, satiety and desire to eat. *European Journal of Clinical Nutrition* **47**, 815–819.

Leathwood P & Pollet P (1988) Effects of slow release carbohydrates in the form of bean flakes on the evolution of hunger and satiety in man. *Appetite* **10**, 1–11.

Lissner L, Levitsky D, Strupp B, Kalkwarf H & Roe D (1987) Dietary fat and the regulation of energy intake in human subject. *American Journal of Clinical Nutrition* **46**, 886–892.

Mattes R (1990) Hunger ratings are not a valid proxy measure of reported food intake in humans. *Appetite* **15**, 103–113.

Mayer J (1955) The regulation of energy intake and body weight. *Annals of the New York Academy of Sciences* **63**, 15–43.

Monello MF & Mayer J (1967) Hunger and satiety sensations in men, women, boys and girls. *American Journal of Clinical Nutrition* **20**, 253–261.

Ohnhaus EE & Alder R (1975) Methodological problems in the measurement of pain: a comparison between the verbal rating scale and the visual analogue scale. *Pain* **1**, 379–384.

Pi-Sunyer X, Kissileff HR, Thornton J & Smith GP (1982) C-terminal octopeptide of cholescystokinin decreases food intake in obese men. *Physiology and Behaviour* **29**, 627–630.

Raben A (1995) Appetite and carbohydrate metabolism. PhD Thesis, Royal Veterinary and Agricultural University, Copenhagen.

Raben A, Tagliabue A & Astrup A (1995) The reproducibility of subjective appetite scores. *British Journal of Nutrition* **73**, 517–530.

Reid CA, Harbron CG, Blundell JE & Stubbs RJ (1998) What are psychometric assessments of appetite asking: a preliminary multivariate analysis. *International Journal of Obesity* **22**, Suppl. 3, 151A.

Rogers PJ, Carlyle J, Hill AJ & Blundell JE (1988) Uncoupling sweet taste and calories: comparison of the effects of glucose and three intense sweeteners on hunger and food intake. *Physiology and Behaviour* **43**, 547–552.

Rolls BJ, Hetherington M & Laster LJ (1988) Comparison of the effects of aspartame and sucrose on appetite and food intake. *Appetite* **11**, 62–67.

Silverstone T & Goodall E (1986) Measurement of hunger and food intake. In *Disorders of Eating Behaviour, a Psychoneuroendocrine Approach*, pp. 129–134 [E Ferrari, editor]. Oxford: Pergamon Press.

Speigel TA, Stunkard AJ, Shrager EE, O'Brian CP, Morrison MF & Stellar E (1987) Effect of naltrexone on food intake, hunger and satiety in obese men. *Physiology and Behaviour* **40**, 135–141.

Spitzer L & Rodin J (1981) Human eating behaviour: a critical review of studies in normal weight and overweight individuals. *Appetite* **2**, 293–329.

Stratton RJ, Stubbs RJ & Elia M (1998*a*) Appetite and food intake during artificial nutrition. *Proceedings of the Nutrition Society* **57**, 96A.

Stratton RJ, Stubbs RJ, Hughes DA, King N, Blundell JE & Elia M (1998*b*) Comparison of the traditional paper visual analogue scales questionnaires with an Apple Newton electronic rating system (EARS) in free living subjects feeding *ad libitum*. *European Journal of Clinical Nutrition* **52**, 737–741.

Stubbs RJ (1995) Macronutrient effects on appetite. *International Journal of Obesity* **19**, Suppl. 5, S11–S19.

Stubbs RJ, Harbron CG, Murgatroyd PR & Prentice AM (1995) Covert manipulation of dietary fat and energy density: effect on substrate flux and food intake in men feeding *ad libitum*. *American Journal of Clinical Nutrition* **62**, 316–330.

Stubbs RJ, Johnstone AM, Harbron CG & Reid C (1998*a*) Covert manipulation of the mixed energy density diets: effect on *ad libitum* food intake in ''pseudo free-living'' humans. *International Journal of Obesity* **22**, 980–987.

Stubbs RJ, Johnstone AM, O'Reilly LM & Poppitt SD (1998*b*) Methodological issues relating to the measurement of food, energy and nutrient intake in human laboratory-based studies. *Proceedings of the Nutrition Society* **57**, 357–372.

Stubbs RJ, O'Reilly LM, Johnstone AM, Franklin MF & Reid C (1998*c*) Description and evaluation of an experimental model to examine changes in qualitative and quantitative feeding behaviour in humans. *European Journal of Clinical Nutrition* **53**, 13–21.

Stubbs RJ, Rowley E, Hughes DA, Johnstone AM, Reid CA, King N, Blundell J, Elia M & Stratton RJ (1997) Evaluating a new electronic appetite rating system (EARS). *International Journal of Obesity* **21**, Suppl. 2, 391.

Stubbs RJ, Van Wyk MCW, Johnstone AM & Harbron C (1996) Breakfasts high in protein, fat or carbohydrate: effect on within-day appetite and energy balance. *European Journal of Clinical Nutrition* **50**, 409–417.

Thompson DA & Campbell RG (1977) Hunger in humans induced by 2 deoxy-D-glucose glucoprivic control of taste preference and food intake. *Science* **198**, 1065–1068.

Thompson DA & Campbell RG (1978) Experimental hunger in man: Behavioural and metabolic correlates of intracellular glucopenia. In *Central Mechanisms of Anorectic Drugs*, pp. 437–445 [S Garrattini and R Samanin, editors]. New York, NY: Raven Press.

Thompson DA, Welle Sl, Lilavivat U, Penicaud L & Campbell RG (1982) Opiate receptor blockade in man reduces 2-deoxy-D-glucose induced food intake but not hunger, thirst, and hypothermia. *Life Sciences* **31**, 847–852.

Trenchard E & Silverstone T (1983) Naloxone reduces the food intake of normal human volunteers. *Appetite* **4**, 43–50.

Westrate JA (1992) Effect of nutrients on the regulation of food intake. In *Unilever Research*. Vlaardingen: Unilever.

Wolkowitz OM, Breier A, Doran A, Rubinow D & Berrettini W (1988) Prednisone-induced behavioural and biological changes in medically healthy volunteers. *Psychopharmacology Bulletin* **24**, 492–494.

Yeomans MR, Gray RW, Mitchell CJ & True S (1997) Independent effects of palatability and within meal pauses on intake and appetite ratings in human volunteers. *Appetite* **29**, 61–76.