https://doi.org/10.1016/j.appet.2014.06.025

Reproducibility of subjective appetite ratings and *ad libitum* test meal energy intake in overweight and obese males

Katy M Horner[1], Nuala M Byrne[1], Neil A King[1]

[1]School of Exercise and Nutrition Sciences, and Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia

All Author Email Addresses: Katy M Horner (katy.horner@chp.edu), Nuala M Byrne (n.byrne@qut.edu.au), Neil King (n.king@qut.edu.au).

Corresponding author:

Katy Horner
Division of Weight Management and Wellness
Childrens Hospital of Pittsburgh of UPMC
Pittsburgh
USA 15201

Tel: +1 4126925294
Email: katy.horner@chp.edu

Abstract

Background: To determine whether changes in appetite and energy intake (EI) can be detected and play a role in the effectiveness of interventions, it is necessary to identify their variability under normal conditions. We assessed the reproducibility of subjective appetite ratings and *ad libitum* test meal EI after a standardised pre-load in overweight and obese males. **Methods:** Fifteen overweight and obese males (BMI 30.3±4.9 kg/m^2, aged 34.9±10.6 years) completed two identical test days, 7 days apart. Participants were provided with a standardised fixed breakfast (1676 kJ) and 5 h later an *ad libitum* pasta lunch. An electronic appetite rating system was used to assess subjective ratings before and after the fixed breakfast, and periodically during the postprandial period. EI was assessed at the *ad libitum* lunch meal. Sample size estimates for paired design studies were calculated. **Results:** Appetite ratings demonstrated a consistent oscillating pattern between test days, and were more reproducible for mean postprandial than fasting ratings. The correlation between *ad libitum* EI on the two test days was r=0.78 (P<0.01). Using a paired design and a power of 0.8, a minimum of 12 participants would be needed to detect a 10mm change in 5h postprandial mean ratings and 17 to detect a 500kJ difference in *ad libitum* EI. **Conclusion:** Intra-individual variability of appetite and *ad libitum* test meal EI in overweight and obese males are comparable to previous reports in normal weight adults. Sample size requirements for studies vary depending on the parameter of interest and sensitivity needed.

Keywords: *ad libitum* energy intake; reproducibility; appetite; obesity; test meal; visual analogue scales.

Introduction

Appetite and energy intake (EI) are often measured in the laboratory using visual analogue scales (VAS) and *ad libitum* test meals respectively. To assess whether these methods are sensitive to detect changes in appetite and EI, it is important to determine their **reproducibility**. The reproducibility of subjective appetite ratings (Barkeling et al. 1995; Raben et al. 1995; Stratton et al. 1998; Arvaniti et al. 2000; Flint et al. 2000; Gonzalez et al. 2012) and *ad libitum* EI (Arvaniti et al. 2000; Gregersen et al. 2008; Nair et al. 2009) has been studied in normal weight adults. Overall, it appears that while the reproducibility will vary depending on the parameter reported, VAS demonstrate a good degree of intra-individual reproducibility (see Stubbs et al. (2000) for a comprehensive review). Similarly, EI at an *ad libitum* meal has been shown to be reproducible in normal weight males (Gregersen et al. 2008). Surprisingly, despite being frequently assessed in response to interventions in overweight and obese individuals, little information exists on the reproducibility of these measures in this population.

Given some evidence that gut peptide (Valera Mora et al. 2005) and appetite (Barkeling et al. 1995) responses may vary according to body composition, it is possible that the reproducibility of appetite and EI will be different in overweight and obese individuals. For example, **a range of factors such as the reward value of food, social desirability, eating behavior characteristics (e.g. disinhibition) and even exercise (King 1999) may be more likely to influence appetite ratings** and EI in overweight and obese individuals (Barkeling et al. 1995; Barkeling et al. 2007). Barkeling et al. (1995) compared VAS ratings in normal weight and obese individuals and demonstrated that while hunger sensations were more reproducible in obese men, desire to eat sensations were less reproducible. However, in this study, VAS ratings were only measured immediately before and after the test meal. In the majority of studies, VAS are generally completed before, immediately after a test meal and

then periodically at regular intervals (varying from 15-30min up to hourly) usually for 3-5h, or until the start of the next meal (Blundell et al. 2010). **To the best of our knowledge the reproducibility of postprandial (3-5h) appetite ratings between two separate days has not been previously documented in overweight and obese individuals.**

With regard to *ad libitum* EI, EI at lunch was previously reported to be highly reproducible in eight overweight/obese individuals using the preload paradigm (Lara et al. 2010). However, the *ad libitum* lunch meal was provided at a relatively short interval (90 minutes) after the preload in this study (Lara et al. 2010). The reproducibility could be influenced by the time interval (Rolls et al. 1991). Therefore, there is a need for further understanding of the reproducibility of subjective appetite ratings and *ad libitum* EI in overweight and obese individuals under standardised conditions. This knowledge can be used to inform appropriate sample sizes for the design of studies investigating changes in these parameters in the pathogenesis or treatment of obesity.

The aims of the present study were to (i) determine the reproducibility of VAS for appetite ratings and *ad libitum* EI following a **more typical** inter-meal interval (ii) calculate minimum sample sizes required to detect hypothetical changes in appetite ratings and *ad libitum* EI **in prospective studies**, and (iii) examine relationships amongst the various appetite measures, in overweight and obese males.

Subjects and Methods

Subjects

Fifteen overweight and obese men [BMI 30.3±4.9kg/m^2, percent body fat 32.1±8.0 %, age 34.9±10.6yrs] participated. **Nine were classified as overweight and six obese by BMI.** Eating behaviour was assessed using the Three Factor Eating Questionnaire (Stunkard and Messick 1985). Height was measured without shoes to the nearest 0.5cm and weight to the nearest 0.01kg. Body composition was measured using air displacement plethysmography (Bodpod, Concord, CA). None of the participants had a history of gastrointestinal disease or surgery, significant illness, or were taking any medication known to affect appetite or EI**. All participants indicated they were willing to consume study test meals described on a screening questionnaire.** The study received ethical approval from Queensland University of Technology Research Ethics Committee.

General Design

Each participant participated in two identical test days 7 days apart. Participants were instructed to refrain from strenuous exercise and alcohol during the 24h prior to each test day. Participants were provided with a standardised evening meal (McCain Beef Lasagne (2447kJ) to consume as their main meal) at home prior to the test day and then fasted for 12h overnight. An identical process was followed prior to the second test day, and prior food intake on the test day was checked by diet recall.

Fixed Breakfast

On the test morning participants were provided with a fixed pancake breakfast spread with butter and strawberry jam [1676 kJ (400 kcal); 15g (15%) PRO, 17g (37%) Fat, 48g (48%) CHO)] with 250ml of water. This meal was used to be consistent with the other studies in this

series of studies on gastric emptying. Breath samples were given as described between breakfast and lunch (Horner et al. (In Press)). Participants remained in the laboratory in sedentary activities throughout.

Appetite

Subjective appetite sensations were measured before and after the fixed breakfast, and periodically during the postprandial period using an electronic appetite rating system (Gibbons et al. 2011). Participants were asked to rate sensations of hunger, fullness and desire to eat on 100 mm visual analogue scales, anchored at each end with the statements not at all and extremely. Postprandial area under the curve (AUC) was calculated using the trapezoidal rule, and 5h mean values calculated by averaging the 6 post-breakfast ratings.

Palatability

Six questions concerning sweet, savoury, tasty, pleasant, filling and satisfying ratings of the test meals were assessed on a 100mm scale using an identical electronic appetite rating system (Gibbons et al. 2011) immediately post consumption of the fixed breakfast and *ad libitum* lunch meals.

Ad libitum Energy Intake

Five hours after the fixed breakfast, participants were provided with an *ad libitum* pasta lunch meal (pasta, tomato sauce and cheese; 47% CHO, 35% FAT, 18% PRO, energy content 7.6kJ/g) and water. **The meal was provided in a large serving dish with utensils and a plate and participants were instructed to help themselves to as much as they wished**

until comfortably full. Water intake was recorded and the amount of food consumed was determined by weighing the meal before and after consumption. EI was calculated using the manufacturers' nutrient values.

Statistical Analysis

Data are expressed as mean ±s.d unless otherwise stated. The Bland and Altman method (Bland and Altman 1986), coefficient of repeatability (CR), Pearson correlation coefficients and coefficient of intra-subject variation (CV_{intra}) were calculated to assess the reproducibility and allow comparison with others. The CV_{intra} was calculated as $CV_{intra} = SD_d/(m\sqrt{2})$ where SDd is the standard deviation of the differences between the repeated tests and *m* is the mean of the repeated tests (Deane et al. 2010, Lartigue et al. 1994). The coefficient of repeatability (CR = 2 x SD) for the mean differences between visits 1 and 2 was calculated (Bland and Altman 1986). The CR indicates the absolute variability of the method whereas the CV measures the relative variability (Gregersen et al. 2008). Pearson correlations were used to determine test-retest correlations and to determine relationships between variables. Based on the standard deviations observed in these parameters, sample size calculations to detect a hypothetical treatment effect with 80% power were calculated using Graph Pad StatMate version 2.0 for Mac (GraphPad Software, San Diego, CA, USA). Sample size calculations were undertaken by selecting the "compare two means (paired t test)" option in the software, entering the observed SD of the difference between pairs and the level of significance of 0.05. Statistical analysis was performed using PASW Statistics 18.0 (SPSS Inc., Chicago, IL, USA). Significance was set at $P<.05$.

Results

All participants (n=15) completed all components of the study. TFEQ scores for dietary Restraint, Hunger and Disinhibition were 7±3, 6±3, and 8±3 respectively.

Appetite

Subjective appetite scores for hunger, fullness and desire to eat are shown in Figure 1.

[Figure 1 About Here]

The reproducibility of fasting VAS scores, breakfast intra-meal differences (pre breakfast minus post breakfast rating), 5h postprandial mean ratings and 5h AUC values for hunger, fullness and desire to eat are shown in Table 1. Correlations between ratings on the first and second visit were strongest for mean 5h and 5h AUC ratings, and were weakest for fasting ratings (Table 1). Similarly, CRs were larger for fasting ratings (range 42-61mm) than mean 5h ratings (range 16-24mm). CVs were largest for fasting ratings (34-68%) and lowest for 5h postprandial mean (11-20%) and 5h AUC (11-19%) ratings.

[Table 1 About Here]

Palatability

Palatability ratings for ‘sweet’, ‘savoury’, ‘tasty’, ‘pleasant’, ‘filling’ and ‘satisfying’ for breakfast and lunch meals were similar between test days (Table 2). The mean ratings for ‘taste’, ‘pleasant’, ‘filling’ and ‘satisfying’ were >50mm on the 100mm scales for both the fixed breakfast and *ad libitum* lunch meals.

[Table 2 About Here]

Power Calculations for Subjective Appetite and Palatability Ratings

Based on the standard deviations observed, sample size estimates required to detect a given change in a paired design study were calculated (Table 3). **The sample size estimates indicate that** smaller sample sizes are needed to detect changes in mean postprandial than fasting ratings (Table 3).

[Table 3 About Here]

Ad libitum Energy Intake

EI was not significantly different between the two visits (visit 1: 4095±1068kJ, visit 2: 4572±1639kJ; P=0.10). There was a small increase in water intake at visit 2 (visit 1: 327±114ml, visit 2: 365±82ml; P=0.04). A Bland Altman plot revealed a mean bias of 477kJ and a single outlier of 3442kJ for EI (Figure 2). Including the outlier in calculations, the CR was ±2129kJ and the CV_{intra} 17.4%. The outlier did not have a significant influence on subjective appetite ratings. However, as the difference in EI between visits for the outlier was 2.8 s.d. from the mean, the outlier was removed for the sample size calculations for *ad*

libitum EI. With the outlier removed the mean bias decreased to -265kJ, the CR to ±1408kJ and the CV_{intra} to 11.5%. To detect a 500kJ and 1000kJ difference in a paired design, minimum sample sizes of 17 and 6 are needed, respectively.

[Figure 2 About Here]

The correlation between EI at visits 1 and 2 is shown in Figure 3. When the outlier was removed, the correlation coefficient increased to r=0.91 (R^2=0.82, P<0.001).

[Figure 3 About Here]

Relationships Amongst Variables

Correlation coefficients of subjective appetite ratings, TFEQ scores and *ad libitum* EI are shown in Table 4. When considering the mean of the two visits, postprandial 5h mean ratings of hunger and desire to eat were correlated with mean *ad libitum* EI. Pre-lunch ratings and intra-meal (i.e., pre-post lunch differences), were not correlated with *ad libitum* EI, with the exception of hunger at visit 2 (Table 4). However, the difference in the pre-lunch fullness rating between the two test days was negatively correlated with EI, indicating an increase in pre-lunch fullness from visit 1 to visit 2 was associated with a reduction in EI at the lunch

meal. There were no associations between mean palatability ratings or the difference in palatability ratings between the two visits and *ad libitum* EI (data not shown).

[Table 4 About Here]

Body weight, BMI, or percentage body fat were not significantly associated with *ad libitum* lunch EI (data not shown). However, there were significant relationships between TFEQ Hunger and Disinhibition and mean *ad libitum* EI (Table 4).

Discussion

Food intake methodology is becoming increasingly important in obesity research (Blundell et al. 2009). It follows that claims about changes in appetite or the effects of interventions in overweight or obese people should be based on evidence from studies on this population (Blundell et al. 2010). The present study assists to inform the design and interpretation of studies by providing novel information regarding the reproducibility of subjective appetite ratings and *ad libitum* test meal EI in overweight and obese males. Based on evidence of altered appetite sensations (Barkeling et al. 2007) and gut peptide release (Valera-Mora et al. 2005) in overweight and obese individuals, it could be anticipated that subjective appetite ratings and *ad libitum* EI would demonstrate poor reproducibility in this population. However, the primary findings of the current study were that the reproducibility of appetite ratings and *ad libitum* EI are comparable to previous reports in normal weight adults (Raben et al. 1995; Flint et al. 2000; Gregersen et al. 2008) and that the intra-individual variability varies depending on the parameter of interest.

A consistent oscillating pattern was evident for subjective appetite ratings, similar to observations in other populations (Whybrow et al. 2005). CVs for 5h postprandial mean ratings and 5hAUC ratings ranged between 11 and 20%, consistent with studies in normal weight adults reporting CVs of between 7-24% for 4.5h mean and AUC ratings (Flint et al. 2000). CRs for fastings ratings were larger than previously reported by Flint et al. (2000) in 55 normal weight adults but were similar to those reported by Raben et al. (1995) in a sample of nine lean males. In addition, correlations between appetite ratings on the two test days were weakest for fasting ratings and strongest for postprandial mean 5h and AUC ratings. Others have similarly reported postprandial mean ratings to be more reproducible than fasting ratings in lean adults (Flint et al. 2000; Gonzalez et al. 2012). Flint et al. (2000) highlighted that this is not surprising as the role of a single erroneous rating is reduced when mean ratings

are calculated. While hourly postprandial ratings over 5h in the present study demonstrated good reproducibility, and were comparable to previous reports in lean adults assessed every 30min over 4.5h (Flint et al. 2000), for shorter duration (e.g. ≤2h) appetite assessments more frequent VAS ratings (e.g. every 15-30min) may be important.

The breakfast intra-meal difference (difference between pre and post breakfast rating) was more reproducible (CRs range 41 – 49 mm) than fasting ratings (CRs 42 – 61 mm) alone, but CRs were still large. One explanation may be the fixed energy content of the breakfast (1676kJ) in the present study. It is likely that the variability in intra-meal differences would be reduced if the energy content of the test meal is increased, or the energy content is tailored to BMI or body weight. Collectively, the present findings demonstrate that the reproducibility of subjective appetite ratings varies depending on the appetite parameter of interest, being most reproducible for postprandial mean and AUC ratings and least reproducible for fasting ratings in overweight and obese males.

With regard to *ad libitum* EI, Lara et al. (2010) previously assessed the reproducibility of *ad libitum* EI at a lunch meal served 90 minutes after a control or whey protein preload in overweight and obese individuals. They reported CVs of 4.5 and 11.2%, and a mean difference between visits of -50kJ and -142kJ respectively, for control and whey preloads. In contrast in the present study, the CVs were higher - 17.4 and 11.5%, the CRS ±2129kJ and ±1408kJ and correlation coefficients 0.76 and 0.90, with and without the outlier respectively, indicating a lower reproducibility. One possible explanation is the longer inter-meal interval (5h) in the present study. Indeed, the present findings are similar to Gregersen et al's. (2008) - in normal weight males (without prior diet standardisation) - who used a similar inter-meal interval (4.5h), and reported a CV of 14.5%, CR of ±1.8MJ and r=0.65 for *ad libitum* EI at lunch. The authors concluded the *ad libitum* test meal is a reproducible method in normal weight males. Further, the variability in the present study did not appear to depend on the size

of EI. The present findings therefore add to previous work (Lara et al. 2010), suggesting that with a longer inter-meal interval (5h) *ad libitum* EI is less reproducible than previously reported after a shorter inter-meal interval (90min) (Lara et al. 2010). However, the reproducibility is consistent with previous reports in normal weight males following a similar inter-meal interval (Gregersen et al. 2008) and provides support that an *ad libitum* test meal is a suitable method of assessing changes in EI in overweight and obese males.

Information regarding the reproducibility of appetite and EI is necessary to determine estimates of appropriate sample sizes. The present data supports anecdotal evidence which suggests 20-25 participants are generally sufficient to capture a 10% difference in the mean or AUC appetite ratings in a paired design (Blundell et al. 2010). However, larger sample sizes are needed to assess changes in fasting appetite sensations in overweight and obese males. This is important to consider when both designing and interpreting studies in this population depending on the appetite variable of interest. The significant influence of an outlier on the reproducibility of *ad libitum* EI was evident in the present study and has similarly been documented by others (Venti et al. 2010). This may represent the extreme of intra-individual variability in this outcome measure and is important to consider when assessing individual changes in *ad libitum* EI.

A number of previous studies support the contention that appetite ratings are related to subsequent food intake when other factors are controlled (Blundell et al. 2010). However, in the present study we found no consistent strong significant associations between appetite ratings and *ad libitum* EI. Moreover, intra-individual variability in *ad libitum* EI did not appear to be associated with palatability ratings. The only variable associated with the difference in EI between Visits 1 and 2 was the difference in the pre-lunch rating for fullness. These findings contrast with evidence in normal weight adults of significant associations between a number of appetite ratings and subsequent EI (Flint et al. 2000).

One explanation for the lack of consistent significant relationships between these parameters in the present study is that a variety of other factors may be more likely to influence appetite ratings and EI in overweight and obese individuals. Such factors could include the reward value of food (Finlayson et al. 2009), 'social desirability' or cognitive factors such as Disinhibition (Barkeling et al. 1995; Barkeling et al. 2007). Barkeling et al. (1995) suggested that obese individuals may find it less threatening to describe a great general hunger (assessed by TFEQ) compared to rating hunger on scales prior to an eating occasion and that 'social desirability' to conform to norms may influence subjective appetite ratings in obese individuals.

Other explanations may be that relationships between appetite ratings and EI are highly dependent on the timing of the meal or dependent on the sample size. For example hunger ratings late in the post meal period (e.g. 4-6h) when most of the previous meal has emptied, may have much less association to EI than differences earlier in the post meal period (Blundell et al. 2010). It is possible therefore that if the *ad libitum* meal was served earlier in the postprandial period (e.g at 3h) relationships between appetite and EI (e.g. mean 0-3h and EI at 3h) may have been stronger. Studies comparing relationships between appetite and EI over a range of inter-meal intervals in overweight and obese individuals may yield further information. In addition *ad libitum* EI was not associated with body weight or composition in the present study. Although others have shown similar findings in a small sample of overweight and obese individuals (Venti et al. 2010), in larger sample sizes relationships between *ad libitum* EI and body composition have been demonstrated (Blundell et al. 2011). Furthermore, as VAS ratings are inherently subjective, genuine inter-individual differences in interpretation of the scale could limit the inter-participant reliability (Stubbs et al. 2000) and thus this may be an additional explanation for the observed dissociations between appetite ratings and EI in obese males in the present study.

There are various methodological aspects to this study and reproducibility studies in general which deserve consideration. The best method to represent intra-individual variability remains a matter of debate (Hopkins, 2000). Although ultimately resulting in the same value, there is some variation in the CV_{intra} formula used in reproducibility studies (e.g. $CV_{intra} = \sqrt{}(SUM(test1-test2)^2/(2xpairs))/mean$ (Flint et al. 2000); and $CV_{intra} = SD_d/(m\sqrt{2})$ where m is the mean and SDd is the standard deviation of the differences between tests (Deane et al. 2010; Lartigue et al. 1994)). For determining the reproducibility of VAS ratings some studies have compared the reproducibility of tests using paired t-tests or correlations (Lappalainen et al. 1993; Porrini et al. 1995). However, neither of these statistical procedures sufficiently describes the reproducibility of a method (Bland and Altman, 1986). As highlighted by Flint et al. (2000), it is clear that a strong correlation is not necessarily synonymous with a low CR and vice versa, indicating the correlation analysis should not be considered in isolation when assessing reproducibility. Therefore, although we have reported a range of different parameters to assess the reproducibility of the different measures we have focused on discussing primarily those which allow comparison with others.

In the present study, although the 5h inter-meal interval may be considered more typical compared to shorter durations (<2h), it is acknowledged that a 5h interval without snacking is not. While currently the most common protocols in appetite studies involve studying changes over an interval without inter-meal snacking, the influence of an inter-meal snack on the reproducibility of VAS and EI would be of interest in future study. Finally, it is important to acknowledge that healthy adult overweight and obese males were studied to exclude any confounding effects of gender and phase of the menstrual cycle on appetite and EI (Brennan et al., 2009). A comparison of overweight with obese individuals was not undertaken due to the sample size in these subgroups. This is an area relevant to consider in

future investigations, along with studies in other populations including overweight and obese females.

To the best of our knowledge this is the first study to investigate the reproducibility of *ad libitum* EI and subjective appetite ratings over an inter-meal interval in an overweight and obese population. The data show that the reproducibility of subjective appetite ratings and *ad libitum* EI is similar to that reported in lean adults and the reproducibility and hence the sample size required for studies varies depending on the parameter of interest. This knowledge will assist in the interpretation of previous studies and design of future studies that aim to investigate changes in these parameters in the pathogenesis or treatment of obesity.

List of Abbreviations

AUC, area under the curve; CR, coefficient of repeatability; CV, coefficient of variation; EI, energy intake; TFEQ, three factor eating questionnaire; VAS, visual analogue scale.

Conflict of Interest

The authors declare no conflicts of interest.

Authorship

The author’s responsibilities were as follows – KMH, NMB and NAK contributed to the design of the study; KMH collected the data, analysed the data and drafted the manuscript; NMB and NAK contributed to data analysis and critical revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are particularly grateful to the participants in this study. This study was supported by a Queensland University of Technology Postgraduate Research Award (QUTPRA).

References

Arvaniti, K., Richard, D. and Tremblay, A. (2000). Reproducibility of energy and macronutrient intake and related substrate oxidation rates in a buffet-type meal. *British Journal of Nutrition* **83**(5): 489-495.

Barkeling, B., King, N. A., Näslund, E. and Blundell, J. E. (2007). Characterization of obese individuals who claim to detect no relationship between their eating pattern and sensations of hunger or fullness. *International Journal of Obesity* **31**(3): 435-439.

Barkeling, B., Rossner, S. and Sjoberg, A. (1995). Methodological studies on single meal food intake characteristics in normal weight and obese men and women. *International Journal of Obesity and Related Metabolic Disorders* **19**(4): 284-290.

Bland, J. M. and Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **i**: 307-310.

Blundell, J., De Graaf, C., Hulshof, T., et al. (2010). Appetite control: methodological aspects of the evaluation of foods. *Obesity Reviews* **11**(3): 251-270.

Blundell, J. E., Caudwell, P., Gibbons, C., Hopkins, M., Naslund, E., King, N. A. and Finlayson, G. (2011). Body composition and appetite: fat-free mass (but not fat mass or BMI) is positively associated with self-determined meal size and daily energy intake in humans. *British Journal of Nutrition* **107**(3): 445-449.

Blundell, J. E., De Graaf, K., Finlayson, G., Halford, J. C. G., Hetherington, M., King, N. A. and Stubbs, J. (2009). Measuring food intake, hunger, satiety and satiation in the laboratory. *Handbook of Assessment Methods for Obesity and Eating Behaviours*. Allison, D. B. and Baskin, M. L. Newbury Park, CA, Sage**:** 283-325.

Brennan, I.M., Feltrin, K.L., Nair, N.S., et al. (2009). Effects of the phases of the menstrual cycle on gastric emptying, glycemia, plasma GLP-1 and insulin, and energy intake in healthy lean women. *American Journal of Physiology Gastrointestinal and Liver Physiology*, **297**(3): 602-610.

Deane, A.M., Zaknic, A.V., Summers, M.J., et al. (2010). Intrasubject variability of gastric emptying in the critically ill using a stable isotope breath test. *Clinical Nutrition*, **29**(5): 682-6.

Finlayson, G., Bryant, E., Blundell, J. E. and King, N. A. (2009). Acute compensatory eating following exercise is associated with implicit hedonic wanting for food. *Physiology & Behav*iour, **97**:62-7.

Flint, A., Gregersen, N. T., Gluud, L. L., et al. (2007). Associations between postprandial insulin and blood glucose responses, appetite sensations and energy intake in normal weight and overweight individuals: a meta-analysis of test meal studies. *British Journal of Nutrition* **98**(1): 17-25.

Flint, A., Raben, A., Blundell, J. E. and Astrup, A. (2000). Reproducibility, power and validity of visual analogue scales in assessment of appetite sensations in single test meal studies. *International Journal of Obesity and Related Metabolic Disorders* **24**(1): 38-48.

Gibbons, C., Caudwell, P., Finlayson, G., King, N. and Blundell, J. (2011). Validation of a new hand-held electronic data capture method for continuous monitoring of subjective appetite sensations. *International Journal of Behavioral Nutrition and Physical Activity* **8**(1): 57.

Gonzalez, J. T., Veasey, R. C., Rumbold, P. L. and Stevenson, E. J. (2012). Consistency of metabolic responses and appetite sensations under postabsorptive and postprandial conditions. *Appetite* **59**(2): 228-233.

Gregersen, N. T., Flint, A., Bitz, C., Blundell, J. E., Raben, A. and Astrup, A. (2008). Reproducibility and power of ad libitum energy intake assessed by repeated single meals. *American Journal of Clinical Nutrition* **87**(5): 1277-1281.

Hopkins, W.G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, **30**(1): 1-15.

Horner, K. M., Byrne, N. M., Cleghorn, G. J. and King, N. A. Reproducibility of gastric emptying in overweight and obese males. *Clinical Nutrition* (In Press).

King, N.A. (1999). What processes are involved in the appetite response to moderate increases in exercise-induced energy expenditure? *Proceedings of the Nutrition Society* **58**(1): 107-13.

Lara, J., Taylor, M. A. and Macdonald, I. A. (2010). Is ad libitum energy intake in overweight subjects reproducible in laboratory studies using the preload paradigm. *European Journal of Clinical Nutrition* **64**(9): 1028-1031.

Lappalainen, R., Mennen, L., van Weert, L., et al. (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**(11): 815-9.

Lartigue, S., Bizais Y., Bruley des Varannes, S., et al. (1994) *Inter- and intrasubject variability of solid and liquid gastric emptying parameters. Digestive Diseases and Sciences*, **39**(1): 109-115.

Nair, N. S., Brennan, I. M., Little, T. J., et al. (2009). Reproducibility of energy intake, gastric emptying, blood glucose, plasma insulin and cholecystokinin responses in healthy young males. *British Journal of Nutrition* **101**(7): 1094-1102.

Porrini, M.,Crovetti, R., Testolin, G., et al. (1995) Evaluation of Satiety Sensations and Food Intake After Different Preloads. *Appetite*, **25**(1): 17-30.

Raben, A., Tagliabue, A. and Astrup, A. (1995). The reproducibility of subjective appetite scores. *British Journal of Nutrition* **73**(4): 517-530.

Rolls, B. J., Kim, S., McNelis, A. L., Fischman, M. W., Foltin, R. W. and Moran, T. H. (1991). Time course of effects of preloads high in fat or carbohydrate on food intake and hunger ratings in humans. *American Journal of Physiology* **260**(4 Pt 2): 756-763.

Stratton, R. J., Stubbs, R. J., Hughes, D., King, N., Blundell, J. E. and Elia, M. (1998). Comparison of the traditional paper visual analogue scale questionnaire with an Apple Newton electronic appetite rating system (EARS) in free living subjects feeding ad libitum. *European Journal of Clinical Nutrition* **52**(10): 737-741.

Stubbs, R. J., Hughes, D. A., Johnstone, A. M., et al. (2000). 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. *British Journal of Nutrition* **84**(4): 405-415.

Stunkard, A. J. and Messick, S. (1985). The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *Journal of Psychosomatic Research* **29**(1): 71-83.

Valera Mora, M. E., Scarfone, A., Valenza, V., Calvani, M., Greco, A. V., Gasbarrini, G. and Mingrone, G. (2005). Ghrelin Does Not Influence Gastric Emptying in Obese Subjects. *Obesity* **13**(4): 739-744.

Venti, C. A., Votruba, S. B., Franks, P. W., Krakoff, J. and Salbe, A. D. (2010). Reproducibility of ad libitum energy intake with the use of a computerized vending machine system. *American Journal of Clinical Nutrition* **91**(2): 343-348.

Whybrow, S., Stephen, J. R. and Stubbs, R. J. (2005). The evaluation of an electronic visual analogue scale system for appetite and mood. *European Journal of Clinical Nutrition* **60**(4): 558-560.

Figure legends

Figure 1 Mean (± SEM) subjective appetite scores on the two test days (Visit 1: black filled circles; Visit 2: grey open circles). Pre bfast: pre breakfast; post bfast: post breakfast; plus1hr: 1 h post breakfast; plus2hrs: 2 h post breakfast; plus 3hrs: 3h post breakfast; plus 4hrs: 4hpost breakfast; plus 5hrs: 5h post breakfast; post lunch: rating immediately after *ad libitum* lunch meal.

Figure 2 Bland–Altman plot for *ad libitum* energy intake (EI) at the lunch test meal. The difference in EI between visits 1 and 2 (y axis) is plotted against the mean of EI for the two visits (x axis). Solid line indicates mean bias (-477kJ). Dashed lines indicate *95% limits of agreement*. n=15.

Figure 3 Energy intake at visit 2 plotted against energy intake at visit 1 at the *ad libitum* lunch meal. r=0.76, P<0.01. n=15.

Tables

Table 1 Reproducibility of Appetite Ratings at Visits 1 and 2. *n*=15.

Variable	Mean (SD) difference (Visit 1 – Visit 2) (mm)[a]	*95% CI*	*P*-value	*CV* (%)	*CR*	*r*
Fasting Ratings						
Hunger (mm)	-4.6 **(28.6)**	**-20.5, 11.3**	0.54	35	57.2	0.55*
Fullness (mm)	-7.6 **(21.0)**	**-9.7, 4.6**	0.21	68	42.0	0.18
Desire to Eat (mm)	-6.5 **(30.2)**	**-23.3, 10.2**	0.42	34	60.5	0.47
Breakfast Intra-meal Difference						
Hunger (mm)	-6.3 **(24.0)**	**-19.6, 7.0**	0.51	62	48.1	0.59*
Fullness (mm)	0.9 **(24.4)**	**-12.6, 14.5**	0.91	41	48.9	0.44
Desire to Eat (mm)	-7.7 **(20.6)**	**-19.2, 3.7**	0.41	48	41.2	0.67**
Mean 5h Ratings						
Hunger (mm)	1.4 **(8.1)**	**-3.1, 5.9**	0.51	11	16.2	0.92****
Fullness (mm)	-3.3 **(11.8)**	**-9.8, 3.2**	0.30	20	23.5	0.66**
Desire to Eat (mm)	1.1 **(9.9)**	**-4.4, 6.6**	0.72	12	19.8	0.90****
Postprandial 5h AUC Ratings						
Hunger (mm.min)	334 **(2407)**	**-999, 1667**	0.60	11	4814	0.92****
Fullness (mm.min)	-852 **(3469)**	**-2773, 1069**	0.36	19	6939	0.67**
Desire to Eat (mm.min)	368 **(3204)**	**-1407, 2143**	0.77	13	6408	0.88****

Breakfast intra-meal difference: the difference between the post-breakfast and pre-breakfast (fasting) rating.

[a]Mean difference: mean of (visit 1 – visit 2).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Table 2 Reproducibility of Palatability Ratings at Visits 1 and 2. *n*=15

Variable	Mean (SD) difference (Visit 1 – Visit 2) (mm)[a]	*95 % CI*	*P*-value	*CV (%)*	*CR (mm)*	*r*
Breakfast						
Sweet	-2.5 (21.3)	-14.3, 9.4	0.67	27	42.7	0.20
Savoury	-0.8 (20.3)	-12.0, 10.4	0.88	36	40.6	0.72**
Tasty	-6.7 (13.6)	-14.2, 1.0	0.08	17	27.2	0.83***
Pleasant	2.7 (15.1)	-5.8, 11.0	0.52	18	30.2	0.83***
Filling	-7.6 (27.5)	-22.9, 7.6	0.30	33	55.1	0.19
Satisfying	-7.9 (19.8)	-18.8, 3.1	0.15	28	39.5	0.60*
Lunch						
Sweet	-9.1 (19.6)	-20.0, 1.7	0.09	50	39.2	0.79***
Savoury	1.5 (17.0)	-7.9, 10.9	0.74	18	33.9	0.73**
Tasty	-1.8 (16.5)	-10.9, 7.3	0.68	16	33.0	0.62*
Pleasant	-5.3 (18.8)	-15.7, 5.2	0.30	19	37.6	0.60*
Filling	-5.0 (13.3)	-12.4, 2.4	0.17	11	26.7	0.67**
Satisfying	-5.3 (15.3)	-13.8, 3.2	0.20	14	30.7	0.68**

[a] Mean difference: mean of (visit 1 – visit 2).

* p < 0.05, ** p < 0.01, *** p < 0.001.

Table 3 Sample size estimates to detect a given change in VAS[a] ratings in a paired design study (α = 0.05 and 80% power).

	Change		
	10mm	**15mm**	**20mm**
Parameter	*Sample size required*		
Fasting Ratings			
Hunger	65	29	17
Fullness	36	17	10
Desire to eat	73	33	19
Breakfast Intra-meal Difference			
Hunger	47	22	13
Fullness	48	22	13
Desire to eat	35	16	10
Mean 5h Ratings			
Hunger	7	<6	<6
Fullness	12	6	<6
Desire to eat	9	<6	<6
Post-Breakfast			
Sweet	37	17	10
Savoury	32	15	9
Tasty	16	8	<6
Pleasant	19	9	<6
Filling	60	28	16
Satisfying	32	15	9
Post-Lunch			
Sweet	31	15	9
Savoury	24	11	7
Tasty	23	11	7
Pleasant	29	14	8
Filling	15	8	<6
Satisfying	20	10	6

Fasting Ratings: rating prior to breakfast; Mean 5h ratings: mean of ratings between breakfast and lunch; Breakfast Intra-meal difference: the difference between the post-breakfast and pre-breakfast rating (i.e. post breakfast – pre-breakfast); Post-Breakfast: palatability ratings of the test meal assessed post breakfast (1647kj) meal; Post-Lunch: palatability ratings of the test meal assessed post *ad libitum* pasta lunch meal.

[a] VAS: 100mm visual analogue scale using a computerised rating system (Gibbons et al. 2011).

Table 4 Correlation coefficients (r) of appetite and TFEQ scores and *ad libitum* lunch energy intake. *n*=15.

	***Ad libitum* Energy Intake**			
Variable	**Visit 1**	**Visit 2**	***Mean*[a]**	***Difference*[b]**
Pre-lunch				
Hunger (mm)	0.39	0.46	0.44	0.34
Fullness (mm)	-0.31	-0.24	-0.20	-0.61*
Desire to Eat (mm)	0.28	0.27	0.28	0.25
Intra-meal difference				
Hunger (mm)	-0.35	-0.54*	-0.46	-0.44
Fullness (mm)	0.15	0.42	0.27	0.31
Desire to Eat (mm)	0.04	-0.40	-0.21	-0.04
5h Mean				
Hunger (mm)	0.55*	0.46	0.52*	0.36
Fullness (mm)	-0.52*	-0.29	-0.40	-0.47
Desire to Eat (mm)	0.60*	0.41	0.52*	0.25
TFEQ				
Hunger	0.66**	0.31	0.51*	-
Disinhibition	0.49	0.48	0.51*	-
Restraint	-0.10	0.11	0.00	-

Pre-lunch: rating immediately prior to lunch; intra-meal difference: the difference between the pre-lunch and post-lunch rating; 5h mean: mean of ratings between breakfast and lunch; TFEQ: Three Factor Eating Questionnaire.

[a]Mean: Mean of Visits 1 and 2

[b]Difference: the difference between Visit 1 and Visit 2.

* $p \leq 0.05$, ** $p < 0.01$.