

**APPetite: Validation of a smartphone app-based tool for the remote measure of free-living subjective appetite**

Holliday, Adrian; Johnson, Kelsie; Kaiseler, Mariana; Crabtree, Daniel R.

*Published in:*  
British Journal of Nutrition

*Publication date:*  
2021

*The re-use license for this item is:*  
CC BY-NC-ND

*This document version is the:*  
Peer reviewed version

*The final published version is available direct from the publisher website at:*  
[10.1017/S0007114521003512](https://doi.org/10.1017/S0007114521003512)

**Find this output at Hartpury Pure**

*Citation for published version (APA):*  
Holliday, A., Johnson, K., Kaiseler, M., & Crabtree, D. R. (2021). APPetite: Validation of a smartphone app-based tool for the remote measure of free-living subjective appetite. *British Journal of Nutrition*.  
<https://doi.org/10.1017/S0007114521003512>

1 **APPetite: Validation of a smartphone app-based tool for the remote measure of free-living**  
2 **subjective appetite**

3 Adrian Holliday<sup>1,2</sup>, Kelsie O. Johnson<sup>3</sup>, Mariana Kaiseler<sup>2</sup> & Daniel R. Crabtree<sup>4</sup>

4

5 <sup>1</sup>Human Nutrition Research Centre, Population Health Sciences Institute, Newcastle University,  
6 Newcastle upon Tyne, UK

7 <sup>2</sup>Institute for Sport, Physical Activity and Leisure, Leeds Beckett University, Leeds, UK

8 <sup>3</sup>Higher Education Sport, Hartpury University, Hartpury, UK

9 <sup>4</sup>Division of Biomedical Sciences, University of the Highlands and Islands, Old Perth Road, Inverness  
10 IV2 3JH, Scotland, UK

11

12 **Corresponding Author:**

13 Dr Adrian Holliday

14 School of Biomedical, Nutritional and Sports Sciences,

15 Faculty of Medical Sciences,

16 Dame Margaret Barbour Building,

17 Wallace Street,

18 Richardson Road,

19 Newcastle upon Tyne, UK

20 NE2 4DR

21 [adrian.holliday@newcastle.ac.uk](mailto:adrian.holliday@newcastle.ac.uk)

22

23 **Short Title:** APPetite: A mobile app for measuring appetite

24

25 **Keywords:** Hunger, Eating Behaviour, Mobile App, Ecological Momentary Assessment

26

27

28

29

30

31 **Abstract**

32 This study determined the validity, reproducibility and usability of a smartphone app – APPetite – for  
33 the measure of free-living, subjective appetite. Validity was assessed compared with the criterion tool  
34 of pen-and-paper visual analogue scale (VAS) (n=22). Appetite was recorded using APPetite and VAS,  
35 one immediately after the other, upon waking and every hour thereafter for twelve hours. This was  
36 repeated the next day with the order of tool reversed. Agreement between tools was assessed using  
37 Bland-Altman analysis. Reproducibility and usability were assessed in a separate experiment (n=22) of  
38 two trials (APPetite vs. VAS), separated by seven days. Appetite was recorded in duplicate upon waking  
39 and every hour for twelve hours using APPetite or VAS. Agreement between duplicate measures was  
40 assessed using Bland-Altman analysis and coefficient of variation (CV) was compared between tools.  
41 Usability was assessed by comparing compliance and by qualitative evaluation. APPetite demonstrated  
42 good criterion validity with trivial bias of 1.65 units/mm·hr<sup>-1</sup> between APPetite- and VAS-derived AUC  
43 appetite scores. Limits of agreement were within a maximum allowed difference of 10%. However,  
44 proportional bias was observed. APPetite demonstrated high reproducibility, with minimal bias (-0.578  
45 units·hr<sup>-1</sup>) and no difference in CV between APPetite and VAS (1.29±1.42% vs 1.54±2.36%, *p* = 0.64).  
46 Compliance was high with APPetite (92.7±8.0%) and VAS (91.6±20.4%, *p* = 0.81). Ninety percent of  
47 participants preferred APPetite, citing greater accessibility, simplified process and easier/quicker use.  
48 While proportional bias precludes using APPetite and VAS interchangeably, APPetite appears a valid,  
49 reproducible and highly usable tool for measuring free-living appetite in young-to-middle-aged adults.

50

51

52

53

54

55

56

57

58

59

60

61

62

## 63 **Introduction**

64 Subjective appetite is typically assessed using the well-established, valid and reliable visual analogue  
65 scale (VAS) method (Flint et al., 2000; Stubbs et al., 2000). This method usually consists of a set of  
66 questions assessing hunger, fullness/satisfaction, desire to eat and prospective food intake (Blundell et  
67 al., 2010). The question is presented with a 100mm horizontal line scale representing the continuum of  
68 subjective perceptions of these constructs of appetite and anchored at each end with extreme responses.  
69 Participants answer, by making a vertical mark on the horizontal line, representing their current  
70 perception on the continuum. The distance from the left-hand anchor to the vertical mark is measured  
71 and a score, in mm, is generated.

72 The VAS method of subjective appetite is typically completed using pen and paper. While inexpensive  
73 and quick to complete, data processing can be time consuming with a risk of human error, resulting in  
74 the misreporting of behaviour. Although suitable for laboratory and supervised settings, the pen and  
75 paper version of VAS harbours limitations for unsupervised, free-living settings. Adherence to pen and  
76 paper scales and diaries is low (Stone et al., 2002), errors in the completion and timing of measures can  
77 be prevalent (Stratton et al., 1998), and ensuring the pen and paper are always about one's person can  
78 be burdensome. In addition, the use of a pen and paper method for large scale data collection is not  
79 environmentally friendly and in free-living studies, data are usually returned through the posting of  
80 hard-copy VAS, which may result in data loss. The regulation of appetite and eating behaviour is  
81 complex and multifaceted, particularly in a free-living setting with social and environmental influences  
82 and cues, as well as physiological and behavioural determinants. As such, a valid, efficient, affordable  
83 and user-friendly method for the large-scale, free-living assessment of appetite perceptions is sought.

84 Electronic scales for the measure of subjective appetite have been developed to overcome some of these  
85 limitations. Electronic scales have been shown to elicit comparable data to pen and paper methods for  
86 the measure of patient outcomes in clinical settings (Muehlhausen et al., 2015), with high rates of  
87 compliance (Hufford & Shields, 2002). The electronic appetite rating systems EARS I (Delargy et al.,  
88 1996) and EARS II (Gibbons et al. 2011), variations of an electronic VAS and sliding-bar scales, have  
89 been developed for the measure of subjective appetite. Iterations of the EARS I, with differing operating  
90 systems and screen size, proved effective at detecting changes in appetite with differing feeding loads  
91 in a laboratory setting; however, some disagreement in measure with the pen and paper VAS tool was  
92 evident, with a tendency for constrained scores with EARS in some instances (Delargy et al., 1996) and  
93 evidence of higher appetite ratings with EARS in women (Whybrow et al., 2006). When used in a free-  
94 living setting, the EARS demonstrated high test-retest reliability and produced appetite ratings not  
95 different to those of pen and paper VAS (Stratton et al., 1998). However, participants rated a preference  
96 for the pen and paper tool, with it deemed more accessible and easier to use, compared with an  
97 unfamiliar handheld electronic device (Stratton et al., 1998). In contrast, the EARS was perceived easier

108 to use in the study of Whybrow et al., (2006), although participants did find it more time consuming to  
109 use than the pen and paper method. Achieving high user satisfaction is vital for effective and compliant  
110 adoption of mobile technology and applications (Zhang & Adipat, 2009), so a better understanding of  
111 the usability of electronic devices for the measure of free-living appetite is warranted.

112 The EARS II, using questions assessing “hunger”, “fullness” and “desire to eat” and completed by using  
113 a stylus to mark a response on a 84mm, 100 unit horizontal line, has been validated in a laboratory  
114 setting (Gibbons et al., 2011). EARS II appetite scores correlated strongly with pen and paper VAS  
115 scores with controlled dietary manipulation, with Bland and Altman analysis demonstrating very low  
116 bias between measures. Despite the pen and paper method being perceived as easier to use by 55% of  
117 participants, the EARS II was rated the preferred tool (Gibbons et al., 2011). However, the reasons for  
118 this preference were not explored.

119 Despite evident benefits of these electronic systems, there are limitations to their use in free-living  
120 settings and on a large scale. These measures require specific devices and software with limited  
121 accessibility. This means that large-scale data collection is limited, and there remains some participant  
122 burden to collecting data, especially at specific times when appetite may be of particular interest (e.g.,  
123 immediately upon waking, immediately post-exercise, immediately post-feeding, when eating “on-the-  
124 go”). This limitation is somewhat overcome with the wrist-worn PRO-Diary© device, which has been  
125 shown to be a valid tool for monitoring free-living subjective appetite in children (Rumbold, Dodd-  
126 Reynolds & Stevenson, 2013). However, such a device is not widely available and accessible.

127 A widely available, accessible and easy-to-use smartphone application for the measure of subjective  
128 appetite in real time was therefore developed to overcome these limitations. Smartphones are well-  
129 placed to monitor behaviour, given the common habit of carrying them on one’s person at all times.  
130 Using the same questions as the traditional VAS method, and with answers provided using an 11-point  
131 Likert scale, the APPetite application was developed to allow for date and time-stamped measures of  
132 subjective appetite that are immediately relayed to the researcher, allowing for real-time, remote  
133 measures within real-life contexts. Such ecological momentary assessment (EMA) methods – those  
134 obtaining measures of behaviour or perceptions in real-time and in one’s natural setting (Stone &  
135 Shiffman, 1994) – have proved effective for measures of free-living food intake (Costello et al., 2017;  
136 Martin et al., 2012; Rollo et al., 2015), but similar tools for the measure of subjective appetite have not  
137 yet been developed and validated. While the Likert scale of APPetite deviates from the more traditional  
138 ungraded line scale, it has been previously shown that categorical and line scale can produce comparable  
139 data (Jeon, O’Mahony & Kim, 2002) and both are accepted and appropriate approaches for measuring  
140 subjective appetite (Blundell et al., 2010). However, this method is yet to be assessed for validity,  
141 reproducibility and usability.

132 The purpose of this study was to determine the validity, reproducibility and usability of an app-based  
133 tool for the remote measure of subjective appetite in free-living settings. Face validity was assessed by  
134 determining the sensitivity of APPetite to hourly changes in subjective appetite. Concurrent validity  
135 was assessed by determining agreement in subjective appetite scores obtained with APPetite and with  
136 the criterion tool of VAS. To understand user compliance and satisfaction, usability was assessed using  
137 a mixed methods approach.

138

## 139 **Experimental Methods**

### 140 *Study Design*

141 Two experiments were conducted to assess validity, test-retest reproducibility, compliance and  
142 preference of the APPetite smartphone application (compatible with both Apple and Android platforms)  
143 for the measure of subjective appetite perceptions. Experiment 1 was a within-subject, counterbalanced,  
144 cross-over study assessing the face and concurrent validity of APPetite, in comparison with the widely  
145 used, validated, criterion tool of the pen and paper VAS. Experiment 2 was also a within-subject,  
146 counterbalanced, cross-over study assessing test-retest reproducibility and compliance. Participants of  
147 Experiment 2 also completed a qualitative questionnaire to assess preferences of APPetite and VAS.  
148 This design has previously been adopted to assess validity and reproducibility of other appetite rating  
149 systems (Stratton et al., 1998).

150 This study was conducted in accordance with the principles and guidelines laid down in the Declaration  
151 of Helsinki, 2013. All procedures were approved by the Ethics Advisory Committee at Leeds Beckett  
152 University.

153

### 154 *Participants and Enrolment*

155 A convenience sample of participants was recruited predominantly from the West Yorkshire and the  
156 Scottish Highlands regions via word-of-mouth and through email and social media advertisement.  
157 Inclusion criteria were: aged 18-70 years, own and able to access a smartphone and able to complete a  
158 pen and paper questionnaire, able to read English. No incentives were offered for participation.

159 Those willing to partake and meeting the inclusion criteria provided written informed consent either in  
160 person or remotely, via email. At this point, participants provided their age, height and weight. Prior to  
161 the experimental trials, participants were provided with paper copies of VAS for each trial day, clearly  
162 labelled, and sent the link to download the APPetite smartphone app, via either email or WhatsApp.  
163 Written and telephone instructions on how to complete both VAS and APPetite were provided and a

164 test measure using both tools was completed to ensure participant competence and technical  
165 proficiency. Participants were then randomly allocated to Experiment 1 or Experiment 2.

166

167 *Experiment 1 – Validity*

168 Participants completed two 12-hour trials on consecutive days. Upon waking, participants completed a  
169 measure of subjective appetite perceptions using both APPetite and VAS tools, one immediately after  
170 the other. This was repeated hourly for 12 hours. In one trial, the APPetite measure was completed first,  
171 followed immediately by the VAS measure, with this order reversed in the other trial. Participants were  
172 encouraged to consider the repeat measure as a separate measure, and not to simply copy their first  
173 measure. The order of the trials was counterbalanced across participants. Participants were encouraged  
174 to set hourly reminders (on a separate application or device, as this function was not available on the  
175 APPetite app) to ensure compliance. Throughout the trial days, participants were encouraged to  
176 consume their habitual diet.

177

178 *Experiment 2 – Test-retest Reproducibility and Usability*

179 Participants completed two 12-hour trials, separated by 7 days. The protocol was similar to Experiment  
180 1; on one trial, two measures of APPetite were completed, one immediately after the other, hourly for  
181 12 hours, from the point the waking. On the other trial, two measures of VAS were completed, one  
182 immediately after the other, hourly for 12 hours, from the point the waking. Participants were  
183 encouraged to consider the repeat measure as a separate measure, and not to simply copy their first  
184 measure. The order of the trials was counterbalanced across participants. Participants were encouraged  
185 to set hourly reminders (on a separate application or device, as this function was not available on the  
186 APPetite app) to ensure compliance. As data was received by the researcher in real-time, missed or late  
187 measures using APPetite were identified. If a measure was late by five minutes, a text reminder was  
188 sent to the participant. If measures were late by >15 minutes, this was deemed a missed or non-  
189 compliant measure. Throughout the trial days, participants were encouraged to consume their habitual  
190 diet.

191 On completion of trial two, participants were provided a link to an online survey to evaluate satisfaction  
192 with the app (see Appendix 1). This included two closed and three open questions. The closed questions  
193 were: “Which method did you find easier to use?”; “If you were going to undertake the study again  
194 what method would you prefer to use?”. Both questions allowed participants to select the following  
195 answers: APPetitite smartphone; pen and pencil; none. The three open questions were: (i) reasons for  
196 preferred choice, (ii) advantages of the APPetite compared to the pen and pencil method; (iii)  
197 disadvantages of the APPetite compared to the pen and pencil method.

198 *Measures of Subjective Appetite Perceptions*

199 Subjective appetite perceptions were measured using VAS and APPetite. Both consisted of four items  
200 relating to four constructs of appetite (“How hungry are you?”, “How full are you?”, “How strong is  
201 your desire to eat?” and “How much would you expect to eat right now?”). These are validated,  
202 commonly used questions for the VAS method of measuring subjective appetite (Flint et al., 2000;  
203 Blundell et al., 2010). The VAS method uses an ungraded 100mm horizontal line, anchored on either  
204 end by extreme answers to the question. The participant answers the question by making a vertical mark  
205 on the horizontal line, representing their feeling on the continuum. This is completed with a pen, on  
206 paper. The score, in mm, is obtained by measuring the distance from the left-hand side anchor. The  
207 participant was asked to note the exact time of recording each measure.

208 The APPetite application uses the same four items. The question is answered using a 11-point Likert  
209 scale (0-10), anchored with the same extreme answers as the VAS. The participant selects the answer  
210 by tapping the screen of their smartphone. The exact time of the measure was automatically recorded.  
211 The data from APPetite is automatically and instantly transferred to a Google Sheets document of the  
212 principle investigator. The APPetite interface can be seen in Figure 1.

213 For both VAS and APPetite, a single composite appetite score was calculated from the four items as of  
214 Holliday & Blannin (2017) and adapted from the 150mm scale of that study for the 100mm scale of the  
215 present study. This was calculated as hunger score + (100-fullness score) + desire to eat score + expected  
216 intake score for VAS, and hunger score + (10-fullness score) + desire to eat score + expected intake for  
217 APPetite. The composite score for APPetite was multiplied by 10, giving a score out of a maximum of  
218 100, for data analysis and direct comparison with VAS score.

219  
220 (Figure 1 here)

221  
222 *Data Analysis*

223 *Validity*

224 The Bland Altman test (Bland & Altman, 1986) was used to assess agreement between APPetite and  
225 VAS scores for Experiment 1. Bias and limits of agreement (LOA), with 95% confidence intervals (CI)  
226 (Stöckl et al., 2004), were calculated. Standardised mean bias was calculated as bias divided by SD of  
227 the criterion (VAS) measure (Hopkins et al., 2009), and interpreted according to the Cohen scale  
228 (Cohen, 1988). A difference or change in VAS appetite score of 10mm (10%) is accepted as a  
229 “reasonable and realistic difference” (Flint et al., 2000); therefore, a value of  $\leq \pm 10\text{mm}/\text{units}$  was set as  
230 the *a priori* maximum allowed difference (Stöckl et al., 2004). For Bland Altman analyses, area under



231 the curve (AUC) values, calculated using the trapezoid method, were used. AUC was calculated  
232 separately for the two experimental days and summated. Regression analysis was also used to provide  
233 further indication of agreement (correlation and standard error of the estimate) and for visual  
234 representation of agreement between raw values. Difference in appetite profiles obtained from APPetite  
235 and VAS was assessed using 2 x 12 factorial analysis of variance (ANOVA) with repeated measures.

236

#### 237 *Test-retest Reproducibility*

238 The Bland Altman test (Bland & Altman, 1986) was used to assess agreement between test-retest  
239 measures for Experiment 2. The AUC, bias, limits of agreement, standardised mean bias and maximum  
240 allowed difference were calculated and interpreted as described above. Regression analysis was also  
241 used to provide further indication of agreement (correlation and standard error of the estimate) and for  
242 visual representation of agreement between raw values. Agreement between pairs of measures were  
243 also assessed by calculating coefficient of variation (CV). The mean CV across the recording period  
244 was then calculated for each participant, with mean CV values compared between APPetite and VAS  
245 tools using a paired samples t-test.

246

#### 247 *Usability*

248 Compliance of measure for Experiment 2 was compared using a paired samples t-test. Data obtained  
249 from quantitative question of the evaluation questionnaire were tallied and presented as frequencies.  
250 Participants' open-ended responses to the survey were analysed using content analyses, acknowledging  
251 its recognized usefulness for health research (Nandy & Sarvela 1997), and a general inductive approach  
252 was used (Bryman & Burgess, 1994). Answers were read several times to identify themes and  
253 categories. All responses were coded by the first and third authors independently into label categories  
254 to increase trustworthiness. The authors agreed on >80% of emerging categories and during critical  
255 discussions established consensus and resolution on all responses coded.

256 A sample size calculation was conducted for Bland-Altman analysis of agreement (Lu et al., 2016).  
257 Based on the mean difference between EARS I and pen-and-paper VAS scores and standard deviation  
258 of the differences of the study of Stratton et al. (1998), a maximum allowed difference of 10mm/units,  
259 and an  $\alpha$  level of 0.05 and a power of 0.8, a sample size of 20 was required.

260 Throughout, data are presented as means  $\pm$  SD in text and as means  $\pm$  SEM in figures. Where relevant,  
261 for t-tests, effect size was calculated as Cohen's d ( $d$ ), with 95% confidence intervals expressed. An  
262 effect size of 0.2 or greater was considered small, 0.5 or greater considered medium and 0.8 or greater

263 considered large (Cohen, 1988). For ANOVA, effect size was calculated as partial eta squared ( $\eta^2_p$ ).  
264 Data was analysed using Statistical Package for Social Science (SPSS, Chicago, IL).

265

## 266 **Results**

### 267 *Participant Characteristics*

#### 268 *Experiment 1*

269 Twenty-six participants were enrolled and allocated to Experiment 1. Twenty-two participants  
270 completed the study (6 men, 16 women; age =  $36 \pm 15$  yrs; height =  $1.69 \pm 0.10$ m; weight =  $66.5 \pm 14.8$ kg;  
271 BMI =  $23.1 \pm 3.4$  kg·m<sup>-2</sup>; 18-24.9 kg·m<sup>-2</sup>, n=16; 25-29.9 kg·m<sup>-2</sup>, n=5; 30-34.9 kg·m<sup>-2</sup>, n=1). Two  
272 participants failed to complete data collection and withdrew, while two were excluded due to  
273 insufficient data (<90% of measures obtained; for those included,  $98.1 \pm 2.7\%$  of measures were  
274 obtained).

#### 275 *Experiment 2*

276 Twenty-six participants were enrolled and allocated to Experiment 2. Twenty-two participants  
277 completed the study (7 men, 15 women; age =  $32 \pm 12$  yrs; height =  $1.71 \pm 0.12$ m; weight =  $70.0 \pm 18.1$ kg;  
278 BMI =  $23.6 \pm 4.1$  kg·m<sup>-2</sup>; 18-24.9 kg·m<sup>-2</sup>, n=15; 25-29.9 kg·m<sup>-2</sup>, n=5; 30-34.9 kg·m<sup>-2</sup>, n=2). Four  
279 participants failed to complete data collection and withdrew from the study.

280

### 281 *Validity*

282 Three participants mistakenly omitted the final measure of each day (obtaining 12 measures, rather than  
283 13 measures over a 12-hour period). To avoid loss of data or extensive missing data analysis, data for  
284 an 11-hour data collection period was analysed for all participants.

285 Appetite profiles as measured by APPetite and VAS are show in Figure 2. There was no difference in  
286 appetite profiles produced by the two tools (measure x time interaction:  $F(23,483) = 1.008$ ,  $p = 0.45$ ,  
287  $\eta^2_p = 0.046$ ).

288

289 (Figure 2 here)

290

291 The AUC values for the total two-day (22-hour) recording period obtained by APPetite and VAS  
292 correlated strongly and significantly ( $r = 0.980$  (95% CI = 0.865 – 0.997),  $p < 0.001$ ,  $\beta = 0.889$  (95%  
293 CI = 0.808 – 0.969), intercept = 6.324 (95% CI = 2.825 – 9.823), SEE = 2.476; Figure 3), but did differ  
294 significantly ( $43.6 \pm 11.0$  vs.  $41.9 \pm 12.1$  units/mm·hour<sup>-1</sup>,  $t(21) = 3.018$ ,  $p = 0.007$ ,  $d = 0.665$ ). Bland-

295 Altman plot for AUC values is shown in Figure 4. Mean bias was  $-1.654 \text{ units/mm}\cdot\text{hr}^{-1}$  (95% CI = -  
296  $2.764 - -0.514 \text{ units/mm}\cdot\text{hr}^{-1}$ ), and standardised mean bias was  $-0.151$  (95% CI =  $-0.255 - -0.047$ ),  
297 representing a trivial bias. Upper and lower LOA were  $3.386 \text{ units/mm}\cdot\text{hr}^{-1}$  (95% CI =  $1.521 - 5.250$   
298  $\text{units/mm}\cdot\text{hr}^{-1}$ ) and  $-6.694 \text{ units/mm}\cdot\text{hr}^{-1}$  (95% CI =  $-8.559 - -4.830 \text{ units/mm}\cdot\text{hr}^{-1}$ ), respectively.  
299 Regression analysis revealed a  $\beta$  value of  $0.099$  (95% CI =  $0.005 - 0.193$ ,  $p = 0.04$ ), indicating  
300 proportional bias.

301

302 (Figure 3 here)

303

304 (Figure 4 here)

305

306

307

### 308 *Test-retest Reproducibility*

309 The AUC for the first measure and repeat measure obtained with APPetite correlated strongly and  
310 significantly ( $r = 0.993$  (95% CI =  $0.954 - 0.999$ ),  $p < 0.001$ ,  $\beta = 0.989$  (95% CI =  $0.935 - 1.042$ ),  
311 intercept =  $-0.075$  (95% CI =  $-2.527 - 2.377$ ), SEE =  $1.037$ ; Figure 5). Bland-Altman plots for APPetite  
312 test-retest scores is shown in Figure 6. Mean bias was  $-0.578 \text{ units}\cdot\text{hr}^{-1}$  (95% CI =  $-1.029 - -0.127$   
313  $\text{units}\cdot\text{hr}^{-1}$ ), and standardised mean bias was  $-0.065$  (95% CI =  $-0.117 - -0.014$ ), representing a trivial  
314 bias. Upper and lower LOA were  $1.416 \text{ units}\cdot\text{hr}^{-1}$  (95% CI =  $0.825 - 2.416 \text{ units}\cdot\text{hr}^{-1}$ ) and  $-2.571$   
315  $\text{units}\cdot\text{hr}^{-1}$  (95% CI =  $-3.571 - -1.980 \text{ units}\cdot\text{hr}^{-1}$ ), respectively. Regression analysis revealed a  $\beta$  value of  
316  $-0.003$  (95% CI =  $-0.058 - 0.049$ ,  $p = 0.86$ ), indicating no proportional bias.

317

318 (Figure 5 here)

319

320 (Figure 6 here)

321

322

323 The AUC for the first measure and repeat measure obtained with VAS correlated strongly and  
324 significantly ( $r = 0.974$  (95% CI =  $0.829 - 0.996$ ),  $p < 0.001$ ,  $\beta = 0.987$  (95% CI =  $0.877 - 1.097$ ),  
325 intercept =  $0.738$  (95% CI =  $-4.021 - 5.497$ ), SEE =  $1.883$ ; Figure 7). Bland-Altman plots for VAS  
326 test-retest scores is shown in Figure 8. Mean bias was  $-0.195 \text{ mm}\cdot\text{hr}^{-1}$  (95% CI =  $-1.031 - 0.642$   
327  $\text{mm}\cdot\text{hr}^{-1}$ ), and standardised mean bias was  $0.066$  (95% CI =  $0.014 - 0.117$ ), representing a trivial bias.  
328 Upper and lower LOA were  $3.408 \text{ mm}\cdot\text{hr}^{-1}$  (95% CI =  $2.043 - 4.774 \text{ mm}\cdot\text{hr}^{-1}$ ) and  $-3.797 \text{ mm}\cdot\text{hr}^{-1}$   
329 (95% CI =  $-5.163 - -2.432 \text{ mm}\cdot\text{hr}^{-1}$ ), respectively. Regression analysis revealed a  $\beta$  value of  $-0.014$   
330 (95% CI =  $-0.124 - 0.096$ ,  $p = 0.80$ ), indicating no proportional bias.

331

332 (Figure 7 here)

333

334 (Figure 8 here)

335

336

337 Mean CV, calculated as the mean for each pair of measures across the recording period, for each  
338 participants, did not differ between APPetite and VAS (3.47% vs. 4.66%,  $t(21) = 1.11$ ,  $p = 0.279$ ). Mean  
339 CV for AUC values also did not differ between APPetite and VAS ( $1.29 \pm 1.42\%$  vs  $1.54 \pm 2.36\%$ ,  
340  $t(21) = 0.481$ ,  $p = 0.64$ ).

341

### 342 *Usability*

343 There was no difference in measurement compliance between APPetite and VAS in Experiment 2  
344 ( $92.7 \pm 8.0\%$  vs.  $91.6 \pm 20.4\%$ ,  $t = 0.244$ ,  $p = 0.81$ ).

345 Twenty-one of the twenty-two participants of Experiment 2 completed the measurement tool online  
346 evaluation survey. Eighteen of the twenty-one (85.7%) found the APPetite tool the easiest of the two  
347 tools to use. The other three participants found no difference in ease of use. Nineteen of the twenty-one  
348 (90.4%) participants expressed a preference for APPetite, should they be asked to repeat the data  
349 collection process using just one of the two tools. The other two participants expressed no preference.  
350 In response to the first open question “what are the reasons for preferring the selected method” from  
351 the answers from the 19 participants selecting the APPetite two main categories emerged labelled  
352 *Accessibility and Simplified Process* and *Easy and Quick numerical display*. For *Accessibility and*  
353 *Simplified Process* category answers included “easier when going out to places and completing on the  
354 phone”. Regarding the *Easy and Quick numerical display* an example of raw answers was “preferred a  
355 number scale and easy to use” For the second question “what, if any do you consider to be an advantage  
356 of the APPetite compared to pen and paper?” three main categories emerged; the first two categories  
357 were the same as in the previous question and a new category labelled *Environmental Friendly* emerged,  
358 with answers explicitly stating that APPetite was “environmentally friendly”. For the third question  
359 “what, if any do you consider to be disadvantages of the APPetite compared to pen and paper?” two  
360 main categories emerged including *Visual reminders of completion* and *Connectivity and IT issues*.  
361 *Visual reminders of completion* included answers such as “less visual reminder to record results”.  
362 *Connectivity and IT issues* included raw answers such as “No battery, malfunctions and no internet”.

363

### 364 **Discussion**

365 We have developed a novel smartphone application – APPetite – for the measure of free-living  
366 subjective appetite. This study aimed to determine the validity, test-retest reproducibility and usability

367 of Appetite. Experiment 1 suggests that APPetite is a valid tool for the measure of subjective appetite.  
368 The appetite profiles obtained by APPetite and VAS were not different, with comparable traces of  
369 subjective appetite over time. This suggests that APPetite is sensitive to typical intra-day changes in  
370 subjective appetite and hence indicates suitable face validity (Blundell et al., 2010) for free-living  
371 measures. Bland-Altman analysis revealed trivial bias of just 1.65 units/mm·hr<sup>-1</sup> between APPetite- and  
372 VAS-derived AUC appetite scores. Further, the limits of agreement, and 95% CI, were within the *a*  
373 *priori* maximum allowed difference of 10%, or 10mm. This indicated strong agreement between the  
374 two tools. However, although AUC values correlated very strongly, mean AUC values were  
375 significantly different. Further, Bland-Altman analysis did indicate proportional bias; APPetite appears  
376 to produce greater values than VAS at lower perceived appetite, but lower values than VAS at higher  
377 perceived appetite. As such, while it can be determined with confidence that APPetite does provide a  
378 valid measure of subjective appetite, the two tools – APPetite and pen and paper VAS – should not be  
379 used interchangeably. Similar conclusions were drawn when previous electronic appetite rating systems  
380 were assessed for validity (Gibbons et al., 2011; Holliday et al., 2014; Stratton et al., 1998; Whybrow  
381 et al., 2006).

382 Experiment 2 demonstrated a high degree of test-retest reproducibility and usability with APPetite. Low  
383 CV values and trivial bias values compared favourably with the criterion tool of pen and paper VAS,  
384 which has previously been shown to be a reliable and reproducible tool for measuring subjective  
385 appetite (Flint et al., 2000). Limits of agreement, along with 95% CI were comfortably within the *a*  
386 *priori* maximum allowed difference for both APPetite and VAS tools. It is possible that the numbered  
387 scale of APPetite did facilitate a higher test-retest reproducibility, compared with the ungraded line of  
388 VAS. Repeat measures, in both Experiment 1 and Experiment 2, were obtained immediately after one  
389 another. This practice is common in studies of this nature (Gibbons et al., 2011; Holliday et al., 2014;  
390 Stratton et al., 1998; Whybrow et al., 2006), as is it important for any measures of agreement to measure  
391 the same phenomenon in the exact same conditions (i.e., at the same time). While one might not expect  
392 appetite to vary much with a small delay of, say one minute, in a free-living setting it is possible for  
393 food cues to impact on appetite perceptions almost immediately. However, it is acknowledged that  
394 agreement between measures could be biased by the participants' memory of the measure they have  
395 just provided, despite the efforts of the researchers to ensure measures were independent and not simply  
396 replicated. This is likely of greater threat to the internal validity for the reproducibility of APPetite, than  
397 for the validity in comparison with VAS, due to the numbered scale on APPetite. It is more likely that  
398 a numbered score out of 10 was remembered and replicated, than a placement of a mark on an ungraded  
399 line was remembered and replicated (or translated into a score out of 10 in the case of Experiment 1).  
400 As such, the very high test-retest reproducibility of APPetite should perhaps be interpreted with some  
401 caution, but the methodological approach adopted was deemed the preferred option for assessing  
402 validity.

403 Compliance did not differ between APPetite and VAS, with a high proportion of measures being  
404 successfully obtained with both tools. Compliance values were similar to those seen in the study of  
405 Stone et al., (2002), when administering paper and electronic diaries for the free-living reporting of pain  
406 in chronic pain patients. Previous studies investigating the validity of electronic systems for the measure  
407 of subjective appetite have typically been conducted in laboratory setting, which does not allow for  
408 measures of free-living compliance (Gibbons et al., 2011; Whybrow, Stephen & Stubbs, 2006), while  
409 one free-living study did not report compliance (Stratton et al., 1998). The inclusion of this important  
410 assessment in the current study strengthens the evidence of APPetite proving a pioneering tool of high  
411 usability in a free-living environment.

412 When assessing compliance, it is important to also consider participant dropout and withdrawal. Only  
413 two participants were excluded from Experiment 1 due to low compliance (<90% of measures  
414 obtained). A further two participants did consider the time commitment of providing measures every  
415 hour too burdensome and withdrew, while two participants withdrew without providing a reason. The  
416 EMA approach of APPetite also allowed for the identification of two participants who provided multiple  
417 measures retrospectively at the end of the day, rather than at the desired time points.

418 Despite no difference in compliance, participants expressed a clear preference for using APPetite than  
419 completing the pen and paper VAS. Findings that over 90% of participants would prefer to use APPetite  
420 for any future recording of free-living subjective appetite – for reasons associated with accessibility, a  
421 simplified process, and easy and quick use – support the rationale for developing a tool such as APPetite.  
422 While previously developed electronic rating systems have been perceived easy to use (Whybrow et al.,  
423 2006), the development of APPetite as a smartphone application afforded the additional benefit of  
424 participants having the tool on their person for much of the time. Our qualitative findings suggest that  
425 participants found that advantages of using the tool related with accessibility, easy to use and  
426 environmentally friendly compared to providing answers in pen and paper. This is of interest, as the  
427 pen and paper method was preferred to the EARS I tool for very similar reasons in the study of Stratton  
428 et al. (1998). It seems the smartphone platform, with which people are familiar and which people tend  
429 to carry on their person, overcomes some of the limitations of earlier electronic devices with regards  
430 usability. Indeed, these reasons seem to be very promising factors for usability purposes across time  
431 and context (Trull & Ebner-Priemer, 2014). Regarding potential disadvantages of the APPetite tool,  
432 these seem to be mainly related with reminders for completion, and IT and connectivity issues.  
433 Automated reminders would prove a useful additional function of APPetite; this should be a primary  
434 focus of future development of this, or similar tools.

435 Although an increased number of people in the 21<sup>st</sup> century use mobile phones and have internet  
436 connection, it is important to consider barriers for certain specific populations where digital literacy or  
437 connectivity limitations may be a problem. It is acknowledged that the study cohort of the present study

438 is largely young-to-middle aged women, representing a demographic of low-deprivation from a more  
439 economically developed country. As such, conclusions regarding usability, in particular, should be  
440 limited to similar cohorts. Usability may be compromised for those with limited access to smartphone  
441 devices and internet connection and older adults (>65 years) are less likely to have and adopt to  
442 smartphone use (Choudrie, Pheeraphuttrangkoon & Davari, 2020). However, the simplicity of  
443 APPetite, with few steps required, simple display of numbered scales and clear instructions aid usability  
444 for older adults (Morey et al., 2019). Of the cohort of the present study, two participants (both of whom  
445 complete Experiment 1 and Experiment 2) were aged over 65 years (both 67 years of age). Compliance  
446 was high for both (both 100% in Experiment 1, and 100% and 85% in Experiment 2), suggesting  
447 suitable usability. Nonetheless, future research should assess validity, reproducibility and, in particular,  
448 usability of APPetite in older adults. As such, we recommend that researchers and practitioners using  
449 the APPetite ensure that participants have equal access to, and capability to use the tool (Fortney et al.,  
450 2011).

451 APPetite, as a novel EMA method, may represent a progressive approach to measuring free-living  
452 subjective appetite. Mobile phone-based EMA methods for measuring free-living food intake have  
453 proved valid and reliable (Rangan et al., 2016; Rollo et al., 2015; Martin et al., 2012), exhibiting greater  
454 precision than traditional pen and paper food diaries (Costello et al., 2017). With specific relation to  
455 measuring subjective appetite, there are a number of operational and practical advantages of APPetite,  
456 as an EMA method, for the researcher. The automatic transfer of data reduces researcher burden and  
457 eliminates the risk of error when recording and inputting pen and paper VAS data. The real-time  
458 collection and transfer of the data to the researcher allows for a more cost-effective and time-efficient  
459 data collection, and for closer monitoring of measurements. This real-time tracking allows for prompts  
460 and reminders should measures be missed, late or completed incorrectly (Stratton et al., 1998), and data  
461 is collected “time-stamped”, which affords the research greater confidence in the validity of the data.  
462 In the present study, two participants were excluded due to observing inaccurate completion of data  
463 collection with APPetite that would not have otherwise been detected with the pen and paper VAS tool  
464 (mis-reported timing of measures and apparent retrospective measures). Hence, the collection of  
465 measures of subjective appetite using APPetite is likely to prove preferable for researchers as well as  
466 participants.

467 It is appreciated that for insightful monitoring and understanding of free-living eating behaviour, there  
468 is benefit in obtaining a number of measures, using an “appetite toolkit” (Gibbons et al., 2019),  
469 especially when considering the limitations of measuring free-living energy intake (Blundell et al.,  
470 2010). As such, the smartphone app-based APPetite tool may prove a useful addition to such a toolkit  
471 for researchers. Combining the use of APPetite with a smartphone-based EMA method of dietary  
472 analysis may prove an effective approach for assessing multiple components of free-living eating

473 behaviour. It is worth acknowledging that the current study did not assess the ability of APPetite score  
474 of subjective appetite to predict free-living food intake. VAS score has been shown to be a weak  
475 predictor of food intake (Flint et al., 2000; Sadoul et al., 2014); it would be of interest to determine the  
476 ability of APPetite-derived measures of subjective appetite to predict food intake and other parameters  
477 of eating behaviour in free-living settings.

478 Despite encouraging evidence of validity, reproducibility and usability, there remain areas for  
479 improvement in APPetite. Monitoring compliance in real-time and sending reminders is a time-  
480 consuming process for researchers. An in-built reminder or alarm would reduce researcher burden and  
481 could improve compliance, especially as some participants perceived the VAS to be easier to remember  
482 due to the visual cue of the paper questionnaire. The limitations of this study must also be  
483 acknowledged. As mentioned earlier, the study cohort was predominantly young-to-middle aged, non-  
484 obese women, and recruited from areas of low-deprivation, which limits recommended use to similar  
485 populations at this stage. The BMI measure also relied on accurate self-report of height and weight,  
486 which was necessary given the free-living, remote nature of data collection. The efficacy of APPetite  
487 to predict eating behaviour was not assessed, which at this stage limits the application of APPetite to  
488 assessing subjective appetite. The sample is also somewhat heterogeneous, with regards age, BMI and  
489 gender, which must be acknowledged when considering the external validity of the findings. However,  
490 there are also some pertinent strengths of this study. The two-experiment, mixed methods design  
491 allowed for the rigorous assessment of validity, reproducibility and usability, all of which are important  
492 considerations for a measurement tool. The statistical analyses conducted provide a thorough and  
493 rigorous assessment of agreement between measures, using *a priori* limits of agreement and an *a priori*  
494 sample size calculation to ensure an appropriate sample size. Further, studies of this nature are typically  
495 not conducted in a free-living setting and hence this study affords assessment of APPetite's  
496 effectiveness as well as efficacy as a tool for free-living, remote measures of appetite.

497 In conclusion, the app-based APPetite tool appears a valid, repeatable and preferred tool for measuring  
498 changes in subjective appetite, compared with the criterion tool of the pen and paper VAS. However,  
499 proportional bias between the two measures suggests that the two tools should not be used  
500 interchangeably. These findings promote APPetite as a viable tool to be used by researchers and  
501 practitioners who wish to remotely measure changes in appetite in free-living settings, specifically in a  
502 cohort of predominantly young-to-middle aged, non-obese women in areas of low deprivation and high  
503 access to mobile phone technology. Further research to assess the validity and usability of APPetite in  
504 other cohorts is needed. Nonetheless, the accessibility to such monitoring could help further our  
505 understanding of appetite regulation, modulation and impact on eating behaviour.

506

507 **Acknowledgements**



508 None

509

510 **Funding**

511 This research received no specific grant from any funding agency, commercial or not-for-profit sectors

512

513 **Conflict of Interest**

514 None

515

516 **Author Contributions**

517 AH formulated the research question. AH and DC designed the study. AH, KJ and DC conducted the  
518 study data collection and data processing. AH and MK conducted data analysis. AH, KJ, MK and DC  
519 interpreted the findings. AH and MK wrote the manuscript. KJ and DC edited the manuscript. All  
520 authors approved the final manuscript draft for submission.

521

522 **References**

523 Bland, J.M., & Altman, D.G. 1986. Statistical methods for assessing agreement between two methods  
524 of clinical measurement. *Lancet*, 327(8476), 307-310.

525 Blundell, J., de Graaf, C., Hulshof, T., Jebb, S., Livingstone, B., Lluh, A., Mela, D., Salah, S.,  
526 Schuring, E., van der Knaap, H., & Westerterp, M. 2010. Appetite control: methodological aspects of  
527 the evaluation of foods. *Obesity Reviews*, 11, 251-270.

528 Choudrie, J., Pheeraphuttrangkoon, S., & Davari, S. 2020. The digital divide and older adult  
529 population adoption, use and diffusion of phone: a quantitative study. *Information Systems Frontiers*,  
530 22, 673-695.

531 Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge  
532 Academic.

533 Costello, N., Deighton, K., Dyson, J., McKenna, J., & Jones, B. 2017. Snap-N-Send: A valid and  
534 reliable method for assessing the energy intake of elite adolescent athletes. *European Journal of Sports  
535 Science*, 17(8), 1044-1055.

536 Delargy, H., Lawton, C., Smith, F., King, N., & Blundell, J. 1996. Electronic appetite rating system  
537 (EARS): validation of continuous automated monitoring of motivation to eat. *International Journal of  
538 Obesity*, 20, 104.

539 Fortney, J.C., Burgess, J.F., Bosworth, H.B., Booth, B.M., & Kaboli, P.J. 2011. A re-conceptualization  
540 of access for 21<sup>st</sup> Century healthcare. *Journal of General Internal Medicine*, 26, S639-S647.

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

544 Gibbons, C., Caudwell, P., Finlayson, G., King, N., & Blundell, J. 2011. Validation of a new hand-held  
545 electronic data capture method for continuous monitoring of subjective appetite sensations.  
546 *International Journal of Behavioural Nutrition and Physical Activity*, 8, 57.

547 Gibbons, C., Hopkins, M., Beaulieu, K., Oustric, P., & Blundell, J. 2019. Issues in measuring and  
548 interpreting human appetite (satiety/satiation) and its contribution to obesity. *Current Obesity Reports*,  
549 8, 77-87.

550 Holliday, A., Batey, C., Eves, F.F., & Blannin, A.K. 2014. A novel tool to predict food intake: The  
551 Visual Meal Creator. *Appetite*, 79, 68-75.

552 Holliday, A., & Blannin, A.K. 2017. Very low volume sprint interval exercise suppresses subjective  
553 appetite, lowers acylated ghrelin, and elevates GLP-1 in overweight individuals: A pilot study.  
554 *Nutrients*, 9, 362.

555 Hufford, M., & Shields, A. 2002. Electronic subject diaries: an examination of applications and what  
556 works in the field. *Applied Clinical Trials*, 11, 46-56

557 Jeon, S-Y., O'Mahony, M., & Kim, K-O. 2002. A comparison of category and line scales under various  
558 experimental protocols. *Journal of Sensory Studies*, 19, 49-66.

559 Lu, M.J., Zhong, W.H., Liu, Y.X., Miao, H.Z., Li, Y.C., & Ji, M.H. 2016. Sample size for assessing  
560 agreement between two methods of measurement by Bland-Altman method. *International Journal of*  
561 *Biostatistics*, 12(2), 2015-0039.

562 Martins, C.K., Correa, J.B., Han, H., Allen, H.R., Rood, J.C., Champagne, C.M., Gunturk, B.K., &  
563 Bray, G.A. 2012. Validation of the Remote Food Photography Method (RFPM) for estimating energy  
564 and nutrient intake in near real-time. *Obesity*, 20(4), 891-899.

565 Morey, S.A., Stuck, R.E., Chong, A.W., Barg-Walkow, L.H., Mitzner, T.L., & Rodgers, W.A. 2019.  
566 Mobile health apps: Improving usability for older adult users. *Ergonomics in Design*, 29(4), 4-13.

567 Muehlhausen, W., Doll, H., Quadri, N., Fordham, B., O'Donohoe, P., Dogar, N., & Wild, D.J. 2015.  
568 Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic  
569 review and meta-analysis of studies conducted between 2007 and 2013. *Health and Quality of Life*  
570 *Outcomes*, 13, 167.

571 Nandy, B.R., & Sarvela, P.D. 1997. Content analysis reexamined: A relevant research method for health  
572 education. *American Journal of Health Behavior*, 21, 222–234.

573 Rangan, A., Tieleman, L., Louie, J., Tang, L., Hebden, L., Roy, R., Kay, J., & Allman-Farinelli, M.  
574 2016. Electronic Dietary Intake Assessment (e-DIA): Relative validity of a mobile phone application  
575 to measure intake of food groups. *British Journal of Nutrition*, 115(12), 2219-2226.

576 Rollo, M.E., Ash, S., Lyon-Wall, P., & Russell, A.W. 2015. Evaluation of a mobile phone image-based  
577 dietary assessment methods in adults with type 2 Diabetes. *Nutrients*, 7, 4897-4910.

578 Rumbold, P.L.S., Dodd-Reynolds, C.J., & Stevenson, E. 2013. Agreement between pen and paper  
579 visual analogue scales and a wristwatch-based electronic appetite rating system (PRO-Diary©), for  
580 continuous monitoring of free-living subjective appetite sensations in 7-10 year old children. *Appetite*,  
581 69, 180-185.

582 Sadoul, B., Schuring, E.A.H., Mela, D.J., & Peters, H.P.F. 2014. The relationship between appetite  
583 scores and subsequent energy intake: An analysis based on 23 randomised controlled studies. *Appetite*,  
584 82, 153-159.

585 Sharp, D.B., & Allman-Farinelli, M. 2014. Feasibility and validity of mobile phones to assess dietary  
586 intake. *Nutrition*, 30, 1257-1266.

587 Stöckl, D., Rodríguez, C.D., Van Uytvanghe, K., & Thienpont, L.M. 2004. Interpreting method  
588 comparison studies by use of the Bland-Altman plot: reflecting the importance of sample size by  
589 incorporating confidence limits and predefined error limits in the graphic. *Clinical Chemistry* 50, 2216-  
590 2218.

591 Stone, A., & Shiffman, S. 1994. Ecological momentary assessment (EMA) on behavioral medicine.  
592 *Annals of Behavioral Medicine*, 16(3), 199-202.

593 Stone, A., Shiffman, S., Schwartz, J., Broderick, J., & Hufford, M. 2002. Patient non-compliance with  
594 paper diaries. *British Medical Journal*, 324(7347), 1193-1136.

595 Stratton, R., Stubbs, R., Hughes, D., King, N., Blundell, J., & Elia, M. 1998. Comparison of the  
596 traditional paper visual analogue scale questionnaire with an Apple Newton electronic appetite rating  
597 system (EARS) in free living subjects feeding *ad libitum*. *European Journal of Clinical Nutrition*,  
598 52(10), 737-741.

599 Stubbs, R., Hughes, D., Johnstone, A., Rowley, E., Reid, C., Elia, M., Stratton, R., Delargy, H., King,  
600 N., & Blundell, J. 2000. The use of visual analogue scales to assess motivation to eat in human subjects:  
601 a review of their reliability and validity with an evaluation of new hand-held computerized systems for  
602 temporal tracking of appetite ratings. *British Journal of Nutrition*, 84(04), 405-415.

603 Trull, T.J., & Ebner-Priemer. 2014. The role of ambulatory assessment in psychological science.  
604 *Current Directions in Psychological Sciences*, 23(6), 466-470.

605 Whybrow, S., Stephen, J., & Stubbs, R. 2006. The evaluation of an electronic visual analogue scale  
606 system for appetite and mood. *European Journal of Clinical Nutrition*, 60(4), 558-560.

607 Zhang, D., & Adipat, B. 2005. Challenges, methodologies, and issues in the usability testing of mobile  
608 applications. *International Journal of Human-Computer Interaction*, 18(3), 293-308.

609

## 610 **Figure Legends**

611 **Figure 1** – APPetite smartphone application. a) welcome page; b) questionnaire interface; c) hunger  
612 item of the questionnaire

613

614 **Figure 2** – Appetite profiles (mean±SEM) for Day 1 and Day 2, as measured using APPetite (solid line,  
615 black circles) and VAS (dashed line, white circles).

616

617 **Figure 3.** Correlation between APPetite and VAS AUC scores over the two-day recording period.  
618 Dashed grey line = line of equity ( $y=x$ ). Solid line = regression line ( $y = 0.889x + 6.324$ ).

619

620 **Figure 4.** Bland-Altman plot for APPetite and VAS scores over the two-day recording period. Solid  
621 black line = mean (grey shaded region = 95% CI). Dashed line = upper and lower limits of agreement  
622 (green shaded area represents 95% CI). Red lines = upper and lower maximum allowed difference. Grey  
623 line = regression line.

624

625 **Figure 5.** Correlation between measure 1 and measure 2 APPetite AUC scores. Dashed grey line = line  
626 of equity ( $y = x$ ). Solid line = regression line ( $y = 0.989x - 0.075$ ).

627

628 **Figure 6.** Bland-Altman plot for measure 1 and measure 2 APPetite AUC scores. Solid line = mean  
629 (blue shaded area represents 95% CI). Dashed line = upper and lower limits of agreement (green shaded  
630 area represents 95% CI). Red lines = upper and lower maximum allowed difference. Grey line =  
631 regression line ( $y = -0.003x - 0.374$ ).

632

633 **Figure 7.** Correlation between measure 1 and measure 2 VAS AUC scores. Dashed grey line = line of  
634 equity ( $y = x$ ). Solid line = regression line ( $y = 0.987x + 0.738$ ).

635

636 **Figure 8.** Bland-Altman plot for measure 1 and measure 2 VAS AUC scores. Solid line = mean (blue  
637 shaded area represents 95% CI). Dashed line = upper and lower limits of agreement (green shaded area  
638 represents 95% CI). Red lines = upper and lower maximum allowed difference. Grey line = regression  
639 line ( $y = -0.014x + 0.384$ ).

640

## 641 **Appendices**

### 642 **Appendix 1 – Method Evaluation Survey**

643

#### 644 **METHOD EVALUATION**

645 Please think back to both methods used to measure appetite and answer the following questions:

646 \*Required

647 This questionnaire is part of the study, titled “APPetite: Validation of an app-based method for the  
648 remote measure of free-living subjective appetite”.

649 1. Do you acknowledge that you have previously provided informed consent to take part in  
650 the study? \*

651

652  Yes, I wish to continue

653

654 2. Please provide a four letter code of the first and last letters of your mother's first name  
655 and maiden name. (For example, if your mother's maiden name is Sarah Johnson, the  
656 code would be "SHJN"). This code will be used to identify your data should you wish  
657 to withdraw from the study. \*

658

659

660 3. What is your age?

661

662 4. If you know it (in either metric or imperial units), what is your height?

663

664

665

666

5. If you know it (in either metric or imperial units), what is your weight?

667

668

669

670

671

672

673

6. What method did you find easiest to use?

674

675

Pen and paper visual analogue scale

676

677

APPetite smartphone app

678

679

I found them equally easy to use

680

681

I found both difficult to use

682

683

684

7. What are the reasons for your answer to Question 6?

685

686

687

8. What, if any, would you consider to be the advantages of the APPetite app, compared with the pen and paper visual analogue scales?

688

689

690

9. What, if any, would you consider to be the disadvantages of the APPetite app, compared with the pen and paper visual analogue scales?

691

692

693

694 10. If you were to take part in a similar study again – recording your appetite throughout the day –  
695 which of the two methods would you prefer to use?

696

697  APPetite app

698

699  Pen and paper visual analogue scale

700

701  I would have no preference

702