Computerized self-administered measures of mood and appetite for older adults: the Novel Assessment of Nutrition and Ageing (NANA) toolkit.

Laura J.E. Brown [a], Tim Adlam [b], Faustina Hwang [c], Hassan Khadra [d], Linda M. Maclean [e], Bridey Rudd [f], Tom Smith [g], Claire Timon [h], Elizabeth A. Williams [i], & Arlene J. Astell [j, k].

[a] School of Psychological Sciences and Manchester Centre for Health Psychology, The University of Manchester, Room S32, Second Floor, Zochonis Building, Brunswick Street, Manchester, M13 9PL, United Kingdom. Email: laura.brown@manchester.ac.uk

[b] Designability, Wolfson Centre, Royal United Hospital, Bath, BA1 3NG, United Kingdom. Email: timadlam@designability.org.uk

[c] School of Systems Engineering, University of Reading, Whiteknights, Reading, RG6 6AY, United Kingdom. Email: f.hwang@reading.ac.uk

[d] Oxford Technologies Limited, 7 Nuffield Way, Abingdon, OX14 1RL, United Kingdom. Email: H.Khadra@gmail.com

[e] Institute of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, Admin building Gartnavel Royal Hospital, 1055 Great Western Road, Glasgow, G12 0XH, United Kingdom. Email: Linda.Maclean@glasgow.ac.uk

[f] School of Social and Health Sciences, University of Abertay, United Kingdom. Email: bridey.rudd@penumbra.org.uk

[g] Generic Robotics Ltd, Reading, United Kingdom. tom@smith.net

[h] Institute of Food and Health, University College Dublin, Room S2.58, Science Centre South, Belfield, Dublin 4, Ireland. Email: claire.timon@ucd.ie

[i] Human Nutrition Unit, Department of Oncology, Faculty of Medicine, Dentistry & Health, University of Sheffield, Sheffield, S10 2RX, UK. Email: e.a.williams@sheffield.ac.uk

[j] Centre for Assistive Technology & Connected Healthcare (CATCH), University of Sheffield, Innovation Centre, 217 Portobello, Sheffield, S1 4DP, United Kingdom. Email: a.astell@sheffield.ac.uk

[k] Ontario Shores Centre for Mental Health Sciences, 700 Gordon Street, Whitby, ON, Canada, L1N 5S9

Corresponding Author: Dr Laura Brown, School of Psychological Sciences, The University of Manchester, Room S32, Second Floor, Zochonis Building, Brunswick Street, Manchester M13 9PL +44 (0)161 275 2563. laura.brown@manchester.ac.uk

Abstract:

The 'Novel Assessment of Nutrition and Ageing' (NANA) toolkit is a computerized system for collecting longitudinal information about older adults' health and behavior. Here, we describe the validation of six items for measuring older adult's self-reported mood and appetite as part of the NANA system. In Study 1, 48 community-living older adults (aged 65-89 years) completed NANA measures of their current mood and appetite alongside standard paper measures, on three occasions, in a laboratory setting. In Study 2, 40 community-living older adults (aged 64-88 years) completed daily NANA measures of momentary mood and appetite in their own homes, unsupervised, alongside additional measures of health and behavior, over three 7-day periods. The NANA measures were significantly correlated with standard measures of mood and appetite, and showed stability over time. They show utility for tracking mood and appetite longitudinally, and for better understanding links with other aspects of health and behavior.

Introduction

Worldwide increases in age-related illness, disease and disability resulting from population ageing have been described as a global health challenge (Esser and Ward, 2013). Reliable and cost-effective ways of assessing and monitoring health status and health behaviors in community-dwelling older adults can help to alleviate this burden this in a number of ways. First, they can facilitate the early detection and treatment of health conditions which, in many cases, can reduce the overall prevalence, impact and severity of the condition (Elia *et al.*, 2005; Fong *et al.*, 2009). Second, home-based assessment methods allow population health needs to be more accurately measured, which can help with the distribution and future planning of health resources. Third, they enable the causes and moderators of poor health to be researched and better understood. And fourth, they are essential for the evaluation of interventions to improve the health, wellbeing and quality of life of older members of society.

Most assessments of health involve collecting data over brief periods of time, and thus provide a single snapshot of the individual's current status. However, in many cases, it is beneficial to collect health information longitudinally, over periods of days or weeks, in order to identify meaningful patterns or changes in health status or behavior (Kaye *et al.*, 2011). Additional benefits can be gained from tracking multiple indices of health simultaneously. In particular, it becomes possible to identify temporal relationships between health variables and behaviors, and thus identify potential causal mechanisms, and develop more effective interventions targeted to individual circumstances (Kang *et al.*, 2010). The availability of cost-effective means of assessing multiple domains of health and behavior over time, that older adults can use in their own homes, would facilitate this.

Patterns of mood and appetite in later life have been linked to a variety of health conditions, such as depression (Engel *et al.*, 2011) and cognitive decline (Kaspar *et al.*, 2015), as well as health behaviors such as physical activity (Powell *et al.,* 2009) and dietary intake (Ford *et al,* 2013). For instance,

changes in appetite caused by depressed mood have been associated with reductions in food consumption and body weight in later life that are known as anorexia of aging (de Boer *et al.*, 2013) . Mood is also an important component of subjective wellbeing (Diener, 2000) and quality of life (Kelley-Gillespie, 2009). Patterns of mood and appetite therefore make useful additions to longitudinal assessments of health and behavior. As both mood and appetite are subjective states that can naturally fluctuate over relatively short timescales (of minutes or hours), they are particularly well-suited to 'momentary' assessment methods that capture patterns of experience by asking respondents to report how they feel at the moment of assessment (Shiffman *et al.*, 2008). Momentary assessment methods have been successfully used with older people (Cain *et al.*, 2009), and help avoid the biases and distortions that are associated with retrospective recall of subjective states (Brown and Astell, 2012). This is particularly important when working with older populations, who may have reduced memory function (Salthouse, 2010).

Digital technology can overcome many of the limitations associated with traditional assessment techniques. As such, we developed the 'Novel Assessment of Nutrition and Ageing' (NANA) toolkit, which is a digital system for capturing momentary information from older adults about their mood and appetite, as well as dietary intake, physical activity, cognitive function and grip strength, over extended periods of time (Astell *et al.*, 2014). The system therefore allows patterns of mood and appetite to be considered alongside these other domains of health and behavior. The NANA system includes a touchscreen computer running custom software that older adults use to enter information about these health and behavioral domains. The system has been designed collaboratively with older adults for unsupervised use, in their own homes, in order to provide a cost-effective method of collecting longitudinal information. Here we describe the development and evaluation of six single-item self-report measures that are used in the NANA system to measure longitudinal patterns of mood and appetite. In Study 1, we determined the reliability and concurrent validity of the computerized items against established pen and paper measures of mood and appetite in a supervised laboratory setting. In Study 2, we investigated the reliability and feasibility of the computerized NANA items

when used by older adults, unsupervised, over extended periods in their own home, as part of the broader NANA toolkit.

Study 1

Methods

Participants

Forty-eight community-living adults (17 males) aged 65-89 years (mean age: 71.6 years) participated in this study. All participants provided written informed consent prior to taking part in the study, which had been approved by the Fife and Forth Valley Committee on Medical Research Ethics (Ref: 08/S0501/104) and the University of St Andrews Teaching and Research Ethics Committee.

Procedure

Each participant was invited to attend three separate, individual testing sessions, over a weeklong period. One participant completed two sessions on the same day (one in the morning and one in the afternoon). All other participants completed their testing sessions on separate days. The mean time difference between consecutive session was 36.5 hours (SD = 20.6), with a range of 4 – 122 hours. All sessions were conducted in an experimental testing laboratory at the University of St Andrews, Scotland.

In the first testing session, participants were given the opportunity to familiarize themselves with a touchscreen computer by performing some standardized practice operations, such as reading text on screen, and making touchscreen responses. Participants then completed a series of touchscreen and pen and paper assessments of cognitive function that formed part of a larger development study for the NANA toolkit. Descriptive measures of global cognitive status, verbal intelligence, and depressive symptoms were also collected using the Mini Mental Status Examination (Folstein *et al.* 1975), National Adult Reading Test (Nelson, 1982), and Geriatric Depression Scale (Yesavage *et al.*, 1983),

6

respectively. After this, computerized and pen and paper assessments of mood and appetite were conducted. The order in which participants completed the computerized and pen and paper measures was counterbalanced between participants.

The mood and appetite assessments were completed again on the following two testing sessions. Two participants were not able to attend the final testing session, and so only contributed data to the first and second testing sessions. All participants received a commemorative mug at the end of their first session, and also received a £5 (approximately $7.5) expenses payment for each session they completed.

Pen and paper measures of mood and appetite

Established pen and paper measures of mood were taken using 39 items from the expanded version of the Positive and Negative Affect Schedule (PANAS-X: Watson and Clark, 1994). This is a reliable scale that has been validated as a measure of two general dimensions of 'positive' and 'negative' affect, as well as eleven specific affective states. The 39 items used were those that comprised the two general dimensions of positive affect and negative affect (10 items each), as well as additional items used to measure five specific affective states of joviality (8 items), sadness (5 items), fatigue (4 items), serenity (3 items), and attentiveness (4 items). The 39 items were presented over five sides of 210 x 297mm paper (7/8 items per page). Instructions presented at the start of the scale asked participants to read each word or phrase, and then to mark the appropriate answer underneath the word in order to indicate the extent to which they felt that way '*right now, that is, in the present moment*'. Each item then comprised of the mood word/phrase printed in black 16 point Times New Roman font, with a 5 point Likert-type response scale underneath. The response options were: 1 (very slightly or not at all), 2 (a little), 3 (moderately), 4 (quite a bit), 5 (extremely), so that higher scores always indicated higher levels of the mood experience. Factor scores for each of the two general dimensions and five specific affective states outlined above were calculated by summing the scores for the items that made up each factor.

Appetite was measured using six visual analogue scale items developed for this purpose (Stubbs *et al.*, 2000). Each item was printed on one page of a 148 x 210mm paper booklet, and comprised of an appetite-related question relating to one of the following six feelings: hunger; fullness; desire to eat; perception of amount they could eat; urge to eat; or preoccupation with thoughts of food. Underneath each question was a 100mm horizontal line, with two opposing statements anchoring the ends of the line. For instance, the question 'how hungry do you feel right now?' had the response anchors of 'not hungry at all' and 'as hungry as I've ever felt'. The instructions printed at the beginning of the booklet asked participants to '*place a mark on each line according to how you feel right now'*. Responses were scored by measuring the distance (in mm) from the anchor to the center of the participant's mark. Scores were reversed where necessary so that higher scores always indicated greater reported appetite. A general 'appetite' factor score was then calculated by summing responses to all six items.

Computerized (NANA) measures of mood and appetite

Computerized mood and appetite items were administered on a 15" touchscreen (Asus EeeTop, model ET1610PT) computer. A written instruction to 'Touch the numbers to show how you feel right now' was presented on the computer screen at the start of assessments, along with a touchscreen button containing the word 'Start' that was used to initiate the first trial. For each trial, one of seven probe words was presented in capital letters in the question string 'How XXX do you feel right now?'. The seven probe words (happy, sad, tired, relaxed, calm, alert, and hungry) were selected on the basis of being commonly used in existing self-report measures of mood or appetite, and for having high face validity as measures of the broad conceptual domains of: happiness, sadness, tiredness, anxiety (two words used: relaxed and calm), alertness and hunger. All seven probe words were also individual items within the PANAS-X or visual analogue scale, allowing concurrent validity to be determined at the item level.

An 11-point (0-10) Likert-type response scale was presented beneath each question string. The scale was anchored with the phrases 'Not at all' and 'Very', which were presented underneath the numbers 0 and 10 respectively. An additional response box containing the words 'No answer' was also

presented in the bottom center of the screen in order to give participants the option of not making a rating for any item. Providing this 'No answer' option thus allowed us to differentiate occasions where participants were engaged with the activity but felt unwilling/unable to provide a response, from those where participants made no active response because they were not engaged with the task or had trouble understanding what to do. Each question string was presented until a touchscreen response was detected, or for 60 seconds if no response was made. This was repeated until all seven probe words had been presented. The order in which the probe words were presented was randomly determined for each testing session. A researcher (B.R.) remained present during the administration of the computerized assessments, but minimized the amount of additional direction or explanation provided during the session.

Results

Participant characteristics

A procedural error meant that two participants were presented with incorrect stimuli during the study, and so their data were excluded from the analysis. The remaining 46 participants (30 female, 16 male) had good levels of general cognitive ability, and low levels of depressive symptoms (Table 1). All participants reported that they had used a computer before. However, levels of computer competence were mixed, with 7 participants (15%) describing themselves as feeling 'not at all' competent at using a computer without assistance; 11 (24%) as 'a little' competent; 23 (50%) as 'fairly' competent; and five (11%) as 'very' competent.

[Table 1 near here]

Reliability of mood and appetite measures

All participants made a valid touchscreen response to all of the computerized mood and appetite items on all occasions tested. A 'no answer' response was made on just two of these occasions (one for the

item 'tired' in the first testing session, and one for the item 'happy' in the second testing session),
comprising 0.2% of data. In contrast, 62 items on the pen and paper scales (1% of the data) were not
completed. In most cases this was because an entire page of items had been missed by the participant.
In addition to these missing data, the paper copy of one participant's appetite responses from the third
testing session was lost before the data had been extracted, and so could not be included in the
analyses.

Participants' responses to several items on both the computerized and pen and paper measures were
not normally distributed, and so non-parametric tests were used for all statistical analyses. As there
were several tied values within the datasets, Kendall's tau tests were used to examine correlations
rather than Spearman's Rank (Field, 2013). Correlation values produced by Kendall's tau tests tend to
be lower than those of Spearman's Rank as they measure different aspects of the bivariate relationship
(Fredricks & Nelsen, 2007).

As mood and appetite are variables that naturally fluctuate over time, the reliability of mood and
appetite measures cannot be precisely determined through simple test-retest correlations.
Nevertheless, individuals do show characteristic 'trait' patterns of mood (Lykken and Tellegen, 1996),
and so some degree of consistency in individuals' scores over time would be expected if measures
were reliable. Cross-session correlations were therefore calculated for each of the NANA items as
well as the corresponding items on the pen and paper PANAS-X and visual analogue scales so that the
degree of consistency between the two measurement types could be compared.

There were significant between-session correlations for all of the NANA and pen and paper items.
The correlation coefficients ranged from .31-.65 for the NANA items, and .30-.82 for the
corresponding pen and paper items (Table 2). For mood, cross-session correlation values were
generally lower for the NANA items than the pen and paper (PANAS-X) items, perhaps due to the
larger range of response options available for the NANA items, which provide greater opportunity for
response variation.

Friedman test of difference revealed significant changes in responses to NANA hungry items across sessions (Table 2). This was due to significantly higher hunger ratings in the first session compared to the second (Wilcoxon signed ranks test: $Z = 2.23$, $p = .03$, two-tailed) and third sessions ($Z = 2.52$, $p = .01$, two-tailed). The same trend was observed for the pen and paper hungry item, but did not reach statistical significance. The higher hunger ratings in session 1 most likely reflect the longer length of the first testing session compared with the second and third, and suggest that the NANA tools are sensitive to short-term changes in appetite.

Interestingly, there was also a significant increase in responses to the pen and paper PANAS calm item from the first to the third testing session (Wilcoxon signed ranks test: $Z = 2.64$, $p = .008$, two-tailed), that was not observed in the NANA items. This may reflect a general increase in calmness with performing pen and paper assessments that did not occur when performing computerized assessments. No significant changes were observed for any of the other mood words. This is consistent with reliable repeated measurement of a variable that shows intra-individual variability, but for which no consistent inter-individual patterns are expected.

[Table 2 near here]

Validity of mood and appetite measures

Concurrent validity was assessed by examining correlations between responses to the computerized mood and hunger assessment items and their corresponding items and factors on the pen and paper PANAS-X and visual analogue scales. Corresponding items in each case were the same word (e.g. the item 'happy' in the PANAS-X scale was used to validate the NANA item 'happy'). Corresponding PANAS-X factors for each NANA mood item were: joviality (for the NANA item 'happy'), sadness (sad), fatigue (tired), serenity (used for the two NANA items 'relaxed' and 'calm'), and attentiveness (alert). The appetite factor taken from the combined responses to the six visual analogue scale items

was used to validate the NANA 'hungry' item. In each case, the score used for each participant was the mean score across the testing sessions for which all corresponding data points were available. Missing scores for the PANAS-X and appetite factor scales were replaced with the participant's average score for the factor from that session.

Significant correlations were found between responses to all of the individual computerized items and the corresponding items and factors on the pen and paper measures. As expected, correlations between individual computerized mood items and the PANAS-X general dimensions of positive and negative affect were more moderate, reflecting the less specific nature of these mood concepts (Watson and Clark, 1994). Data for all of these correlations are displayed in Table 3.

[Table 3 near here]

Discussion

The low level of missing data, the consistency of scores of time, and the high levels of concurrent validity for the computerized items indicate that they offer valid and reliable means of measuring participants' current mood and appetite. The significant correlations between the computerized items and the pen and paper factor scores also indicate that the single items used are good exemplars of the mood and appetite factors selected. Concurrent validity for the probe word 'calm' was slightly lower than for the word 'relaxed', and so the 'relaxed' item was taken as the best indicator of the affective domain of anxiety/serenity. The probe word 'calm' was not retained for Study 2.

Study 2

Having studied responses to the NANA measures in a laboratory setting, the aim of Study 2 was to examine whether similar levels of usability and reliability could be established when participants completed the items in their own homes, unsupervised, as part of the broader NANA toolkit.

Methods

Participants

Forty community-living adults (24 female, 16 male) aged 64-88 years (mean age: 71.7 years) gave written informed consent to participate in this study. Half of the participants were recruited (by CT) around the area of Sheffield, England, and the other half were recruited (by LMM) around the area of St Andrews, Scotland. All participants provided written informed consent prior to taking part in the study, which had been approved by the Fife and Forth Valley Committee on Medical Research Ethics (Ref: 08/S0501/104) and the University of St Andrews Teaching and Research Ethics Committee.

Procedure

Mood and appetite data were collected as part of a larger validation study of the entire NANA toolkit, which is reported in detail elsewhere (Astell *et al.*, 2014). Briefly, each participant was asked to use a NANA system in their home for three separate periods, each approximately one-week in duration, and spaced over three months. Each system's hardware comprised a 15" Asus Eeetop touchscreen computer (model ET1610PT), a webcam (for photographing food and drink items), and a digital hand dynamometer (for measuring grip strength). The software on the computer allowed participants to record their dietary intake, cognitive ability, and levels of activity and exhaustion, as well as their mood and appetite, during the 7-day measurement periods. It also provided instructions for recording grip strength. Participants also underwent additional examination of their physical abilities, dietary intake, and nutritional, cognitive and health status (including blood and urine analysis) through assessments conducted during home and clinic visits before and after the NANA measurement periods. On completion of the study, participants received a payment of £100 (approximately $150) for their involvement in the study.

Participants were shown how to use the system by a researcher who installed it in their homes. During this initial visit, participants practiced performing the various tasks until they felt comfortable using

the system. They were also provided with a simple manual explaining how to use the system, and were given a telephone number that they could use if they required additional support during the assessment period.

Participants were asked to use the system to record their entire dietary intake during each measurement period. All non-dietary assessments were scheduled to be administered on particular days at particular times. Mood and appetite assessments were scheduled to be administered once per day at 10:30am, and were followed immediately by cognitive and physical activity assessments. When one or more non-dietary assessment (such as the mood and appetite assessments) was due to be completed, a message stating that *'There are readings or exercises overdue. Do you want to do them now?'* would be displayed on the screen when participants next interacted with the system, along with an icon that indicated how many were due. Participants could either select "Yes" to complete the assessments at that time, or they could select "No" to postpone them until later. Postponed assessments would continue to show on the system until the participant completed them. Postponed assessments, along with any newly-scheduled ones, were administered in the order in which they were scheduled to appear. Where an assessment was not completed before the next one was due, multiple assessments would be administered within the same session.

Computerized 'NANA' measures of mood and appetite

Mood and appetite were assessed using similar measures to those in Study 1. On each trial, one of six probe words (happy, sad, tired, relaxed, alert and hungry) was presented in the question string 'How XXX do you feel right now?'. The same 11-point (0-10) Likert-type response scale, anchored with the phrases 'Not at all' and 'Very', was presented beneath the question string, along with the additional 'No Answer' response box. Each question string was presented for a maximum of 60 seconds, or until a touch screen response was detected. The six probe words were presented in a randomly-determined order on each occasion.

Results

Participant characteristics

Participants had good levels of current cognitive ability, and low levels of depressive symptoms (Table 1). Thirty-six of the participants (90%) reported having previously used a computer, and 32 (80%) reported being Internet users.

Each participant should have completed seven iterations of the mood and appetite assessments in each of the three weekly-periods (i.e. resulting in an expected total of 840 datasets). However, in some cases, the number of datasets collected was greater or less than this, resulting in a total of 832 datasets. Reasons for this included: administrative errors that resulted in the systems being collected too early, or insufficient trials being scheduled; technical errors with the systems that led to additional trials being presented; and participants residing away from home during part of the measurement period. One participant also failed to enter any data during the first weekly session due to a traffic accident. An additional 51 sets of mood and hunger response sets were subsequently excluded from the final analyses: nine of these (1.08% of the total data) were excluded because corresponding cognitive data had not been collected (indicating a possible fault or anomaly with the data collection session), and 42 (5.05% of the total data) were excluded for being collected within 15 minutes of a previous period of data collection (and thus not representing a separate period of assessment). The final analysis therefore related to 781 sets of six mood and hunger trials, taken from 40 participants across the three weekly periods: 248, 264, and 269 from the first, second and third testing week respectively.

Usability of mood and appetite measures

In order to examine usability, the percentage of each of the three response types (i.e. numerical mood ratings, 'no answer' responses, and missing data due to no response being made before the trial timed

out) were calculated for each participant. These were examined separately for each word. That is, percentages of each response type for each participant were first calculated for each word type in each of the weekly sessions. Mean percentages were then calculated across the sessions. Table 4 shows that the percentage of 'no answer' and 'timed out' trials was low overall, indicating that participants were generally able and willing to provide responses to the items. Although participants showed a trend towards making fewer numerical responses made to sad and happy words than other mood words, the difference between word types did not reach statistical significance. Inspection of individual datasets showed that one participant gave 'no answer' responses to all of the 'happy' questions in all three testing sessions, and almost all of the 'sad' questions. In total, this participant contributed 60% of the 'no answer' responses, with the other 40% of 'no answer' responses being made by 12 other participants.

Response types for the 38 participants who contributed data to all three sessions were also compared across sessions. For these analyses, data for each word type were combined. Friedman tests (with exact levels of significance) showed that participants' average percentage of trials in which a mood rating (χ^2=4.00, df=2, p = .15), 'no answer' (χ^2=4.91, df=2, p = .09), or 'timed out' (χ^2=2.00, df=2, p = .56) response occurred did not significantly differ across the three testing weeks.

[Table 4 near here]

Reliability of mood and appetite measures

Correlations between average scores across testing sessions ranged from .42 - .77 for all items. The correlation coefficients are generally higher than those of Study 1 due to the longer period of data collection covered in Study 2. There were also no significant differences in scores for each item across testing sessions (Table 5). As with Study 1, this is consistent with repeated testing of variables that vary at the individual level, but are not expected to show consistent patterns of inter-individual change.

Discussion

The low levels of missing data, and the consistency of mean scores over time, suggest that the NANA items are useable and reliable means of assessing mood and appetite as part of a broader suite of health assessments. Importantly, as participants performed the assessments in their own homes, without a researcher being present, these measures offer an efficient means of assessing mood and appetite longitudinally in community settings.

General discussion

The six computerized NANA mood and appetite questions provided valid, reliable, and acceptable means of gathering self-report data about momentary mood and appetite in community-living older populations. The detailed analyses of participants' behavior showed relatively few occurrences of missed responses, even when participants used the system unsupervised, in their own homes. Taken together with the consistency of average responses seen across testing sessions, and the significant correlations of scores with established single and multi-item measures of mood and appetite, these results indicate that the six NANA items provide an effective means of gathering valid longitudinal data about self-reported mood and appetite as part of a broader battery of health and lifestyle measures. These NANA mood and appetite items are therefore well-suited to examining relationships between mood, appetite and other aspects of health status and behavior, as well as measuring changes in mood and appetite over time, which may be useful for both clinical and research purposes.

Due to their inherently subjective nature, objective 'gold-standard' measures of self-report states, such as mood and appetite, are not available. Instead, construct validity must be inferred through proxy measures, such as their ability to predict relevant behaviors or outcomes (Chronbach and Meehl, 1955). Self-report measures of mood and appetite have been shown to have high predictive validity for associated behaviors and health conditions (Ford *et al*., 2013; Powell *et al.*, 2009; Stubbs *et al*., 2000), and are also responsive to interventions that aim to alter them (Brown *et al.*, 2008; Stubbs *et al.*, 2000). The high levels of concurrent validity between the NANA items and the pen and paper

measures of mood and appetite in Study 1 therefore indicate that the NANA items are also likely to show high predictive and criterion-related validity in health and lifestyle contexts. An essential next step for future research is to now examine the predictive validity of the NANA items when administered in the home, without a researcher present.

One feature of the NANA mood and appetite assessments was the response option that participants could use to indicate that they wished to provide 'no answer' to a particular item. Although participants' reasons for choosing this response were not recorded in the current study, the trend towards more frequent 'no answer' responses when responding to 'sadness' or 'happiness' items suggests that some participants were less willing or able to quantify their feelings in these domains. It is also possible that a change in the frequency of 'no answer' responses may, in itself, be informative about an individual's mood or general health. The predictive validity of 'no answer' responses is therefore a fruitful avenue for future research.

Despite the high validity of the NANA measures, some participants expressed frustration during or after testing sessions that they were not able to alter a response once they had touched one of the on-screen response buttons. This design feature was chosen in order to minimize the number of button presses and time that participants spent on the task, and also to prevent participants from retrospectively editing their responses according to their memories, expectations, or desires for how they were feeling at a previous time point. Nevertheless, given that some participants may occasionally touch response buttons inadvertently, thereby resulting in an unintended and inaccurate response being submitted, future studies should explore the relative pros and cons of enabling participants to immediately correct a response that was knowingly made in error, such as with the option of an 'undo' button.

There are some limitations to these studies. Notably, participants in both studies were relatively high functioning adults, with high levels of education. The majority of participants also had at least some previous experience of using computers. It is therefore not clear how valid , reliable or acceptable the

NANA system would be when used by participants with higher levels of impairment, or in those with lower levels of education or computer experience. Similarly, given the highly visual nature of the NANA items, it is likely that adaptations to the system would be required for participants with significant levels of visual impairment.

Despite these limitations, the results of these two studies provide evidence that single-item, computerized measures of mood and appetite, such as those used in the NANA system, can be used to collect meaningful self-report information from relatively high functioning older participants over extended periods of time. Moreover, these methods can be used by older participants, in their own homes, without the need for supervision from a researcher. They therefore provide an effective means of gathering longitudinal data than can help to identify individuals who may be experiencing changes in their mood or appetite, as well as understanding more about the links between mood, appetite, and other health-related variables and conditions.

Conflicts of Interest: none.

References

Astell, A.J., Hwang, F., Brown, L.J.E., Timon, C., Maclean, L.M., Smith, T., … Williams, E.A. (2014). Validation of the NANA (Novel Assessment of Nutrition and Ageing) touchscreen system for use at home by older adults. *Experimental Gerontology, 60,* 100-107.

Brown, A.K., Liu-Ambrose, T., Tait, R. & Lord, S.R. (2008). The effect of group-based exercise on cognitive performance and mood in seniors residing in intermediate care and self-care retirement facilities: a randomised controlled trial. *British Journal of Sports Medicine, 43,* 608-614. doi: 10.1136/bjsm.2008.049882

Brown, L.J.E. & Astell, A. (2012). Assessing mood in older adults: a conceptual review of methods and approaches. *International Psychogeriatrics, 24,* 1197-1206. doi: 10.1017/S1041610212000075

Cain, A.E., Depp, C.A. & Jeste, D.V. (2008). Ecological momentary assessment in aging research: a critical review. *Journal of Psychiatric Research, 43,* 987-996. doi: 10.1016/j.jpsychires.2009.01.014

Chronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

de Boer, A., Ter Horst, G.J. & Lorist, M.M. (2013). Physiological and psychological age-related changes associated with reduced food intake in older persons. *Ageing Research Reviews, 12,* 316-328.

Diener, E. (2000) Subjective well-being: the science of happiness and a proposal for a National index. *American Psychologist, 55,* 34-43.

Elia, M., Zellipour, L. & Stratton, R.J. (2005). To screen or not to screen for adult malnutrition? *Clinical Nutrition, 24,* 867-884.

Engel, J.H., Siewerdt, F., Jackson, R., Akobundu, U., Wait, C. & Sahyoun, N. (2011). Hardiness, depression, and emotional well-being and their association with appetite in older adults. *Journal of the American Geriatrics Society, 59*, 482-487. doi: 10.1111/j.1532-5415.2010.03274.x.

Esser, D.E. & Ward, P.S. (2013). Ageing as a global public health challenge: from complexity reduction to aid effectiveness. *Global Public Health, 8,* 745-768. doi: 10.1080/17441692.2013.817598

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th edn.) London: Sage.

Fredricks, G.A. & Nelsen, R.B. (2007). On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables. *Journal of Statistical Planning and Inference, 137,* 2143-2150.

Folstein, M.F., Folstein, S.E. & McHugh, P.R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189-98.

Fong, T.G., Tulebaev, S.R. & Inouye, S.K. (2009). Delirium in elderly adults: diagnosis, prevention and treatment. *Nature Reviews Neurology, 5,* 210-220. doi: 10.1038/nrneurol.2009.24

Ford, P.A., Jaceldo-Siegl, K., Lee, J.W., Youngberg, W. & Tonstad, S. (2013). Intake of Mediterranean foods associated with positive affect and low negative affect. *Journal of Psychosomatic Research, 74,* 142-148. doi:10.1016/j.jpsychores.2012.11.002

Kang, H.G., Mahoney, D.F., Hoenig, H., Hirth, V.A., Bonato, P., Hajjar, I. & Lipsitz, L.A. (2010). In situ monitoring of health in older adults: technologies and issues. *Journal of the American Geriatrics Society, 58,* 1579-1586. doi: 10.1111/j.1532-5415.2010.02959.x.

Kaspar, R., Oswald, F., Wahl, H.-W., Voss, E. & Wettstein, M. (2015). Daily mood and out-of-home mobility in older adults. Does cognitive impairment matter? *Journal of Applied Gerontology, 34,* 26-47: doi: 10.1177/0733464812466290.

Kaye, J.A., Maxwell, S.A., Mattek, N., Hayes, T.L., Dodge, H., Pavel, M., … Zitzelberger, T.A. (2011). Intelligent systems for assessing aging changes: home-based, unobtrusive, and continuous assessment of aging. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 66B*, i180-i190. doi: 10.1093/geronb/gbq095

Kelley-Gillespie, N. (2009). An integrated conceptual model of quality of life for older adults based on a synthesis of the literature. *Applied Research in Quality of Life*, *4,* 259–282. doi: 10.1007/s11482-009-9075-9

Lykken, D. and Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science, 7,* 186-189.

Nelson, H.E. (1982). *The National Adult Reading Test (NART): test manual.* NFER-Nelson:Windsor.

Powell, R., Allan, J.L., Johnston, D.W., Gao, C., Johnston, M., Kenardy, J., … Rowley, D.I. (2009). Activity and affect: repeated within-participant assessment in people after joint replacement surgery. *Rehabilitation Psychology, 54,* 83-90.

Salthouse, T. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society, 16,* 754-760. doi: 10.1017/S1355617710000706

Shiffman, S., Stone, A.A. & Hufford, M.R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4,* 1-32.

Stubbs, R.J., Hughs, D.A., Johnstone, A.M., Rowley, E., Reid, C., Elia, M., … Blundell, J.E. (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,* 405-415.

Watson, D. & Clark, L.A. (1994). *The PANAS-X: manual for the positive and negative affect schedule – expanded form.* Iowa City: University of Iowa.

Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M. & Leirer, V.O. (1983). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research*, *17,* 37–49.

Table 1: Characteristics of participants included in Study 1 (*n* = 46), and Study 2 (*n* = 40).

	Study 1 *M* (*SD*), Range	Study 2 *M* (*SD*), Range
Age (years[†])	71.78 (5.90), 65-89	71.70 (5.85), 64-88
MMSE score (/30)	28.11 (1.98), 23-30	28.63 (1.64), 22-30
NART error score (/50)	13.17 (8.53), 2-35	11.40 (6.48), 1-34
GDS score (/15)	1.32 (1.73), 0-7	1.62 (1.91), 0-8

MMSE = Mini Mental Status Examination (Folstein et al., 1975); NART = National Adult Reading Test (Nelson, 1982); GDS = Geriatric Depression Scale (Yesavage et al., 1983). [†] Numerical age, not rounded to nearest age.

Table 2: Mean scores for each of the seven computerized NANA mood and appetite items and their corresponding items on the pen and paper tasks in Study 1. Friedman statistics and Kendall's tau correlation co-efficients refer, respectively, to tests of difference and correlations between scores for each item across testing sessions. Note that scores for NANA tests are out of a maximum of 10; scores for paper (PANAS-X) mood items are out of 5; and scores for the paper hungry item are out of 100.

Item	Task	Session 1 Mean score (*SD*)	Session 2 Mean score (*SD*)	Session 3 Mean score (*SD*)	Fried-man χ^2 (*df*=2)	Cross-session correlation		
						1vs.2	2vs.3	1vs.3
Happy	NANA	7.69 (1.98)	7.67 (1.48)	7.33 (1.79)	2.39, ns	.40**	.54***	.49***
(*n*=42)	Paper	3.64 (0.96)	3.62 (0.88)	3.81 (0.74)	2.24, ns	.51***	.71***	.55***
Sad	Nana	1.29 (2.81)	0.88 (1.95)	0.69 (1.84)	2.38, ns	.47***	.58***	.35**
(*n*=42)	Paper	1.26 (.077)	1.29 (0.81)	1.24 (0.73)	0.64, ns	.62***	.82***	.46**
Tired	Nana	2.83 (3.13)	2.66 (2.97)	2.00 (2.52)	0.83, ns	.32**	.44***	.31**
(*n*=41)	Paper	1.66 (0.83)	1.71 (0.87)	1.76 (0.83)	0.66, ns	.51***	.51***	.30*
Relaxed	Nana	6.91 (2.81)	7.48 (2.39)	7.64 (2.08)	2.69, ns	.44***	.51***	.45***
(*n*=44)	Paper	3.80 (0.88)	3.95 (0.75)	4.00 (0.72)	3.59, ns	.53***	.67***	.47***
Calm	Nana	7.02 (2.77)	7.52 (2.03)	7.48 (2.05)	0.12, ns	.42***	.65***	.47***
(*n*=42)	Paper	3.40 (0.96)	3.64 (1.00)	3.81 (0.77)	6.51*	.45***	.72***	.49***
Alert	Nana	7.38 (2.18)	7.60 (1.52)	7.19 (2.18)	1.17, ns	.32**	.44***	.36**
(*n*=42)	Paper	3.67 (0.69)	3.76 (0.69)	3.69 (0.64)	0.66, ns	.31*	.70***	.51***
Hungry	Nana	2.88 (2.60)	2.12 (2.28)	1.88 (1.98)	9.29**	.41***	.44***	.43**
(*n*=41)	Paper	26.3 (22.3)	16.8 (17.0)	18.0 (17.7)	5.43, ns†	.36**	.57***	.37***

ns = non-significant; * $p < .05$; ** $p < .01$; *** $p < .001$; † $p = .07$ (significance levels calculated for one-tailed hypotheses).

Table 3: Concurrent validity between computerized and pen and paper measures of mood and appetite in Study 1. Kendall's tau correlation co-efficient values refer to correlations between participants' responses to each of the seven computerized mood and appetite words, and the corresponding pen and paper equivalents. Correlations are shown for the pen and paper responses to: individual items; composite factor scores for affective/appetite states (named in brackets); and general positive and negative affect dimension scores.

Item	Correlation with corresponding item on pen and paper scale.	Correlation with corresponding factor on pen and paper scale.	Correlation with PANAS-X Positive Affect dimension.	Correlation with PANAS-X Negative Affect dimension.
Happy	.66***	.55*** (serenity)	.42***	-.24*
Sad	.47***	.38** (sadness)	.04, ns	.28**
Tired	.55***	.59*** (fatigue)	-.25*	.25*
Relaxed	.58***	.52*** (serenity)	.29**	-.22*
Calm	.46***	.50*** (serenity)	.26**	-.26**
Alert	.49***	.55*** (attentiveness)	.44***	-.30**
Hungry	.66***	.58*** (appetite)		

ns = non-significant; * $p < .05$; ** $p < .01$; *** $p < .001$ (significance levels calculated for one-tailed hypotheses).

Table 4: Mean percentage (and range) of response types given to each word in Study 2.

Friedman statistics refer to tests of difference between words for each response type. Asymptotic

significance values are provided.

	Happy Mean (range)	Sad Mean (range)	Tired Mean (range)	Relaxed Mean (range)	Alert Mean (range)	Hungry Mean (range)	Friedman χ^2 (df=5)
% of numerical rating responses	96.89 (0-100)	96.36 (14-100)	98.40 (78-100)	99.05 (89-100)	99.06 (89-100)	99.39 (89-100)	10.78, $p = .06$
% of 'no answer' responses	3.12 (0-100)	3.26 (0-89)	1.31 (0-22)	0.83 (0-11)	0.94 (0-11)	0.61 (0-11)	5.15, $p = .40$
% of 'timed out' trials	0 -	0.26 (0-6)	0.12 (0-5)	0.12 (0-5)	0 -	0 -	5.00, $p = .42$

Table 5: Means of weekly mean scores for each word collected across each session in Study 2.

Friedman and Kendall's tau statistics refer, respectively, to tests of difference and correlations between scores for each item across testing sessions. Note that sample sizes are slightly lower for happy and sad items due to one participant giving almost all 'no answer' responses to these items.

Item	Session 1 Mean (SD) of mean scores	Session 2 Mean (SD) of mean scores	Session 3 Mean (SD) of mean scores	Fried-man χ^2 (df=2)	Cross-session correlations (r_s) 1vs.2	2vs.3	1vs.3
Happy (n=37)	6.98 (1.39)	6.92 (1.46)	6.92 (1.49)	0.45, ns	.53***	.66***	.65***
Sad (n=37)	1.67 (1.65)	1.82 (1.63)	2.13 (1.71)	3.47, ns	.58***	.63***	.58***
Tired (n=38)	3.18 (1.69)	3.22 (1.85)	3.60 (1.85)	4.96, ns	.51***	.66***	.54***
Relaxed (n=38)	6.42 (1.54)	6.56 (1.68)	6.65 (1.75)	0.89, ns	.59***	.77***	.59***
Alert (n=38)	6.85 (1.32)	6.79 (1.54)	6.72 (1.44)	1.60, ns	.55***	.75***	.55***
Hungry (n=38)	3.14 (1.73)	3.23 (1.82)	3.46 (2.10)	1.29, ns	.48***	.54***	.42***

ns=non-significant; *** $p < .001$ (significance levels calculated for one-tailed hypotheses).