Appetite 123 (2018) 352–356

Contents lists available at ScienceDirect

Appetite

journal homepage: www.elsevier.com/locate/appet

Assessment of test-retest reliability of a food choice task among healthy individuals

Karin Foerde [a,*], Loren Gianini [a,b], Yuanjia Wang [c,d], Peng Wu [d], Daphna Shohamy [e], B. Timothy Walsh [a,b], Joanna E. Steinglass [a,b]

[a] New York State Psychiatric Institute, New York, NY, USA
[b] Department of Psychiatry, Columbia University Medical Center, New York, NY, USA
[c] Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY, USA
[d] Division of Biostatistics, New York State Psychiatric Institute, New York, NY, USA
[e] Psychology Department and Kavli Institute for Brain Science, Columbia University, New York, NY, USA

ARTICLE INFO

Article history:
Received 5 May 2017
Received in revised form
18 October 2017
Accepted 7 January 2018
Available online 10 January 2018

Keywords:
Test-retest reliability
Eating disorders
Anorexia nervosa
Obesity
Food choice
Longitudinal

ABSTRACT

Aberrations in eating patterns constitute a substantial public health burden. Computer-based paradigms that measure responses to images of foods are potentially useful tools for assessing food attitudes and characteristics of eating behavior. In particular, food choice tasks attempt to directly probe aspects of individuals' decisions about what to eat. In the Food Choice Task participants rate the healthiness and tastiness of a variety of food items presented one at a time. Next, participants choose for each food item whether they prefer to eat the item vs. a neutrally rated reference food item. The goal of the current study was to assess the stability and reliability of this Food Choice Task over time and with repeated testing. Secondary analyses were conducted using data from healthy volunteers in two separate studies that administered the task at two time points, separated either by several days or about a month. The overall reliability of the Food Choice Task across multiple administrations was assessed using intra-class correlation coefficients and the reliability of ratings of individual food items was assessed using kappa coefficients. The results indicated that test-retest reliability of the Food Choice Task in healthy volunteers was high at both shorter and longer test-retest intervals. In addition, the reliability of individual food item ratings was good for a majority of items. The proportion of healthy volunteers' high-fat food choices did not change over time in either of the two studies. Thus, the Food Choice Task is suitable for measuring food choices in studies with multiple assessment points. In particular, the task may be well suited to assess restrictive eating, a construct which it has been difficult to assess in experimental settings.

1. Introduction

Food is a primary reward. It is a basic necessity and source of pleasure to many, yet a source of distress to others. While much is understood about basic feeding mechanisms, many questions remain—particularly regarding disturbances in eating behavior that may contribute to obesity and eating disorders, which affect large proportions of the population and represent a substantial public health burden. Thus, there is a pressing need to understand the development and persistence of maladaptive eating behavior, such as over-eating or extreme dietary restriction, and the effectiveness of interventions to change such behavior. Computer-based tests of food-related behavior are a potential means of advancing this understanding (e.g., Foerde, Steinglass, Shohamy, & Walsh, 2015; Steinglass, Foerde, Kostro, Shohamy, & Walsh, 2015).

Numerous food-related behavioral tasks exist. Some tasks have focused on hedonics of food (Rangel, 2013), while others have assessed aspects of reward learning or processing (O'Doherty, 2004). Choice tasks using food stimuli more directly investigate how individuals make decisions about what to eat. In one such task, the Food Choice Task (Foerde et al., 2015; Hare, Camerer, & Rangel, 2009; Steinglass et al., 2015), participants rate images of food according to healthiness as well as tastiness. Based on these ratings they are then offered a choice between a food that they consider

* Corresponding author. 1051 Riverside Drive, Unit 98, New York, NY 10032, USA.
E-mail address: kf2265@columbia.edu (K. Foerde).

https://doi.org/10.1016/j.appet.2018.01.010
0195-6663/

"neutral" and a series of other foods. Food choice tasks have the advantage of measuring decision-making around food by directly probing personal preferences. There are no learning requirements or right and wrong answers within the task. Another attractive task feature is that individualized assessments of food along two dimensions (healthiness and tastiness) allow the tasks to be used across populations that may vary greatly in their valuations of food. For example, among individuals with eating disorders, the subjective value of different foods may vary substantially compared with that observed among healthy peers. The Food Choice Task addresses this issue by obtaining ratings of specific foods for each participant.

The original version of the Food Choice Task was developed to examine self-control in healthy populations (Hare et al., 2009) and focused on "junk foods" and "healthy snacks." To assess decision-making in populations with eating disorders over a more representative range of dietary choices, we adapted the task to assess foods with a broader range of caloric density and macronutrient content. In addition, foods are categorized as low vs. high-fat foods, which confers a specific advantage when assessing dietary restriction. Dietary restriction is extreme in anorexia nervosa (AN), and is characterized by a specific avoidance of calories from fat (Hadigan et al., 2000; Mayer, Schebendach, Bodell, Shingleton & Walsh, 2012; Walsh, 2011), operationalized in the task as the proportion of high-fat foods chosen (Steinglass et al., 2015). This Food Choice Task has been shown to capture the dietary restriction seen among individuals with AN (Steinglass et al., 2015). Additionally, this task has been shown to be a valid assessment of actual restriction in dietary intake, as the proportion of high-fat choices on the task was significantly correlated with actual food intake among individuals with AN (Foerde et al., 2015). The relationship between food choices on the task and real food intake suggests that this task may be a useful assessment of real-world eating.

The goal of the current study was to assess the test-retest reliability of the adapted Food Choice Task in healthy individuals in order to determine the stability and reliability of this task over time and with repeated testing. We conducted reliability analyses, using intra-class correlations (ICC), on data from two studies that administered the task at two time points, separated by several days or by about one month. Analyses were conducted using data from healthy volunteers who did not have eating disorders and were not attempting to change weight or eating behavior. We predicted high reliability for food choices on the task. In addition, we predicted no change in the proportion choices of high-fat foods over time.

2. Methods

2.1. Participants

Participants were 36 female, healthy volunteers, recruited for participation across two studies (Study 1 and Study 2). Volunteers were included if they were between the ages 18–45 years, had no current or past psychiatric illness, including any history of an eating disorder, and were normal weight (in Study 2, one individual was overweight and one was mildly obese). Clinical diagnoses were ruled out using both the Structured Clinical Interview for DSM-IV (SCID) (Spitzer, Williams, & Gibbon, 1987) and the Eating Disorders Examination (EDE) (Fairburn & Cooper, 1993), as well as a clinical interview with a doctoral level clinician. Additional exclusion criteria were significant medical illness, current psychotropic medication, or dietary restrictions (such as vegetarianism, or religious restrictions that would impact food choices in the task).

Participants in Study 1 ($n = 15$) were the subset of healthy volunteers who returned for a second testing session, after participating in a study in which they were compared with individuals with AN (Steinglass et al., 2015). Participants in Study 2 ($n = 21$) were recruited to serve as a healthy control group in a study of individuals with bulimia nervosa (Gianini et al., 2016). Both studies were approved by the New York State Psychiatric Institute Institutional Review Board, and all participants provided written informed consent. Clinical characteristics are presented in Table 1. Demographic characteristics did not differ significantly between the two studies.

2.2. Procedures

Participants completed the Food Choice Task (Fig. 1) on two separate days. The task consisted of three phases. In each phase participants were presented with 43 images of food items. The food items represented a range of dietary options (Steinglass et al., 2015). Twenty-five food items were low fat (<30% calories from fat) and 18 were high fat (>30% calories from fat), as determined by our research nutritionist. The inclusion of foods categorized as low vs. high fat was undertaken in order to adapt the task for use in individuals with eating disorders, who consume significantly fewer calories from fat specifically, relative to healthy individuals (Hadigan et al., 2000; Mayer et al., 2012). In the Health phase, participants rated the healthiness of each food item on a 5-point scale, with 1 indicating "Unhealthy" and 5 indicating "Healthy". In the Taste phase, participants rated the tastiness of each food item on a 5-point scale, with 1 indicating "Bad" and 5 indicating "Good". In the Choice phase, participants made a choice on each trial between the presented food item and a "Neutral" reference food item (rated as 3 in both Health and Taste phases). If no item was rated 3 on both scales, an item rated 3 on Health and greater than 3 on the Taste scale was selected as a reference food. This was done to avoid conflict between health and taste ratings for the reference item and to select a reference item that was as neutral as possible for participants making choices based on health information. The reference food did not change and remained visible throughout the Choice phase (an image of the item was presented next to the computer screen). Most participants had different reference foods at Time1 and Time 2, except for three participants in Study 1 and one in Study 2. There was no time limit for responding in any phase.

For the Choice phase in Study 1, participants were instructed to imagine that they would receive one of their selections as a snack after the study. The task was generally conducted in the afternoon and food consumption prior to the task was not standardized. In Study 2, participants received a snack consisting of one of the foods chosen in the task, randomly selected, after the task. Participants received a standardized lunch, and the Food Choice Task was administered two hours later.

In Study 1, testing procedures at Time 1 and Time 2 occurred approximately one month apart (Mean = 35.0 ± 4.8 days, range: 27–43 days) and were identical. In Study 2, testing procedures at Time 1 and Time 2 occurred a few days apart (Mean = 3.0 ± 2.3 days, range: 1–9 days) and differed slightly. Study 2 included a mood (affect) manipulation (in random, counterbalanced order): on one study day, participants wrote down a neutral memory (the route by which they had arrived at the test site), and, on the other

Table 1
Clinical characteristics of participants in Study 1 and Study 2.

	Study 1	Study 2
	Mean ± SD	Mean ± SD
Age (years)	26.5 ± 6.0	26.1 ± 4.8
BMI (kg/m^2)	21.3 ± 1.8	22.4 ± 3.5
EDE-Q, Total Score	0.1 ± 0.14	0.2 ± 0.23

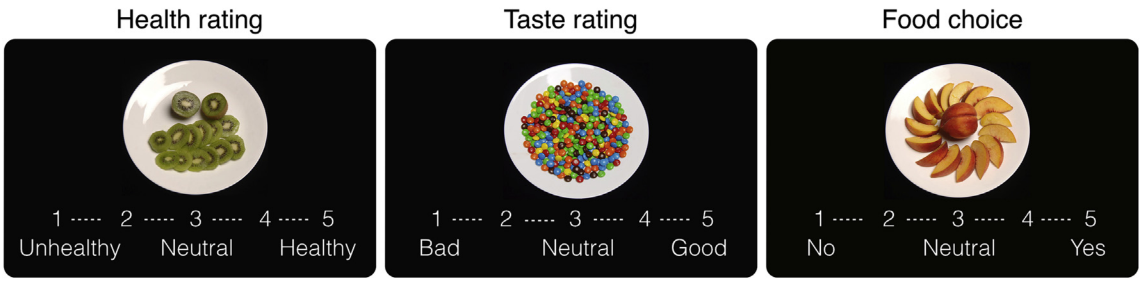

Fig. 1. Food Choice Task. Participants rated 43 foods in three phases. In the Health and Taste phases they rated each food image on a 5-point Likert scale. In the Choice phase, they indicated strength of preference for the food item, as compared with their own individually rated neutral reference item. "No" indicated selection of the reference item, which was visible next to the computer screen, and "Yes" indicated selection of the item on that trial.

study day, participants wrote down a negative memory. This manipulation was intended to induce a negative mood in order to compare food choices during neutral and negative moods. The affect induction did not affect food ratings or choices in healthy volunteers (Gianini et al., 2016) and test-retest reliability was compared between Time 1 and Time 2 (see Results section).

2.3. Statistical analyses

Reliability of the Food Choice Task across multiple administrations was assessed using intra-class correlation coefficients (ICC) for the outcome measures of interest. The ICC(1,1) form of the intraclass correlation was used (Shrout & Fleiss, 1979) for all of the outcome measures. Interpretation of ICC values was as follows: Poor (<0.40), Fair (0.4–0.6), Good (0.6–0.75), and Excellent (0.75–1.0) (Cicchetti, 1994). ICC was computed using the ICC package (Wolak, Fairbairn, & Paulsen, 2012) for R (Team, 2013). To test for significant differences between ICCs for high-vs. low-fat foods, a bootstrap method was used with 1000 iterations to compute the *p* value (Efron & Tibshirani, 1994).

In addition, we assessed whether responses were differentially reliable for specific food items, by calculating kappa coefficients for each food item in each of the three task phases. Where appropriate, a quadratic weighted kappa coefficient was used. When only two levels of ratings were available for an item (e.g., every participant rated an item as either a 4 or a 5), un-weighted kappa was used (Feng & Wen, 2010). Interpretation of kappa values was as follows: Less than chance (<0), Slight (0.01–0.2), Fair (0.21–0.4), Moderate (0.41–0.6), Substantial (0.61–0.8), and Almost perfect (0.81–1.0) (Landis & Koch, 1977). Kappa coefficients were computed using SAS software, version 9.4.

The task outcome measures for the Health and Taste phases were the participants' mean ratings for high-fat and low-fat foods. In the Choice phase, the outcome measures of interest were the individual's percentage of high-fat and low-fat choices of food items over the reference food item. These behavioral data were analyzed using repeated measures ANOVA within the IBM SPSS Statistics 23 analysis package.

3. Results

3.1. Test-retest reliability

The ICCs and confidence intervals for both studies are reported in Table 2.

Study 1: A high degree of reliability was found between measures at Time 1 and Time 2. The ICCs fell in the Excellent range (0.75–1.0), with the following exceptions: healthiness ratings of low-fat foods fell in the Good range (0.6–0.75); Choice phase responses for low-fat foods fell in the Fair range (0.4–0.6). ICCs did not differ significantly between high-fat and low-fat foods in any of the task phases (Health: $p = .99$; Taste: $p = .99$; Choice: $p = .84$).

Study 2: A high degree of reliability was found between measures at Time 1 and Time 2. The ICCs fell in the Excellent range (0.75–1.0), with the following exceptions: healthiness ratings of low-fat foods fell in the Good range (0.6–0.75); tastiness ratings of low-fat foods fell in the Fair range (0.4–0.6); and Choice phase responses for low-fat foods fell in the Poor range (<0.4). ICCs were significantly lower for low-fat foods than high-fat foods in the Choice phase ($p < .001$), but did not differ significantly in the other task phases (Health: $p = .99$; Taste: $p = .65$).

3.2. Food-item reliability

The reliability of ratings of the individual food items across time varied considerably in both Study 1 and Study 2. In all task phases a majority of items were classified, according to the kappa coefficient, as indicating Fair or better agreement (see Table 3) (Landis & Koch, 1977).

3.3. Food task behavior

Study 1: Results are presented in Fig. 2, panels A–C. Behavioral data from each task phase were analyzed in 2 (Time: time 1/time 2) X 2 (Food type: high-fat/low-fat) repeated measures ANOVAs. In the Health phase, high-fat foods were rated as less healthy than low-fat foods overall ($F_{1,14} = 1610.34$, $p < .0001$), with no significant effect of Time ($F_{1,14} = 3.67$, $p = .08$) or interaction with Time ($F_{1,14} = 0.10$, $p = .75$). In the Taste phase there was no difference between ratings of high-fat and low-fat foods ($F_{1,14} = 1.18$, $p = .30$) nor was there a significant effect of Time ($F_{1,14} = 4.28$, $p = .06$) or

Table 2
Intraclass correlation coefficients (ICC) and confidence intervals (CI) for each task phase.

	Task phase	Food type	ICC[a]	Lower CI	Upper CI
Study 1	Health	Low fat	**0.678**	0.399	0.956
		High fat	**0.872**	0.749	0.995
	Taste	Low fat	**0.779**	0.577	0.982
		High fat	**0.890**	0.783	0.997
	Choice	Low fat	**0.416**	−0.009	0.842
		High fat	**0.841**	0.689	0.992
Study 2	Health	Low fat	**0.787**	0.622	0.952
		High fat	**0.654**	0.405	0.902
	Taste	Low fat	**0.592**	0.311	0.873
		High fat	**0.795**	0.636	0.954
	Choice	Low fat	**0.259**	−0.145	0.663
		High fat	**0.738**	0.541	0.935

[a] Classification of ICC values: Poor (<0.40), Fair (0.4–0.6), Good (0.6–0.75), Excellent (0.75–1.0) (Cicchetti, 1994).

Table 3
Number of food items with kappa values falling within each level of kappa magnitude[a] for each task phase.

	Task phase	Almost perfect (0.81–1) agreement	Substantial (0.61–0.8) agreement	Moderate (0.41–0.6) agreement	Fair (0.21–0.4) agreement	Slight (0.01–0.2) agreement	Less than chance (<0) agreement
Study 1	Health	4	9	12	8	3	7
	Taste	5	14	15	3	5	1
	Choice	4	9	16	10	3	1
Study 2	Health	6	12	12	7	1	5
	Taste	8	19	13	3	0	0
	Choice	3	13	16	11	0	0

[a] Kappa magnitude classified according to (Landis & Koch, 1977).

interaction with Time ($F_{1,14} = 0.016$, $p = .90$). In the Food Choice phase, there was a significant main effect of Food type, such that participants chose high-fat foods over the reference item less often than they chose low-fat foods over the reference item ($F_{1,14} = 7.30$, $p = .02$). There was no significant effect of Time ($F_{1,14} = 0.87$, $p = .37$) or interaction with Time ($F_{1,14} = 0.89$, $p = .36$) indicating that choices were similar across Time 1 and Time 2.

Study 2: First we assessed whether the affect manipulation (see Methods) had an effect on results. Behavioral data from each task phase were analyzed in 2 (Affect manipulation: neutral/negative) X 2 (Food type: high-fat/low-fat) repeated measures ANOVAs. There were no main effects of the Affect manipulation (Health phase: $F_{1,20} = 0.20$, $p = .66$; Taste phase: $F_{1,20} = 0.51$, $p = .48$; Choice phase: $F_{1,20} = 0.26$, $p = .62$), nor any Affect manipulation X Food type interactions (Health phase: $F_{1,20} = 2.55$, $p = .13$; Taste phase: $F_{1,20} = 0.43$, $p = .52$; Choice phase: $F_{1,20} = 0.74$, $p = .40$). As negative affect did not influence task behavior in healthy participants, we analyzed behavioral data in 2 (Time: time 1/time 2) X 2 (Food type: high-fat/low-fat) repeated measures ANOVAs (as for Study 1). Results over time are presented in Fig. 2, panels D–F. In the Health phase, high-fat foods were rated as less healthy than low-fat foods overall ($F_{1,20} = 1053.69$, $p < .0001$), with no significant effect of Time ($F_{1,20} = 1.87$, $p = .19$) or interaction with Time ($F_{1,20} = 0.96$, $p = .34$). In the Taste phase, there was no difference between ratings of high-fat and low-fat foods ($F_{1,20} = 1.13$, $p = .30$) nor was there a significant effect of Time ($F_{1,20} = 2.97$, $p = .10$) or interaction with Time ($F_{1,20} = 0.072$, $p = .79$). In the Food Choice phase, there was no significant effect of Food type ($F_{1,20} = 3.57$, $p = .07$) and no significant effect of Time ($F_{1,20} = 0.09$, $p = .77$) or interaction with Time ($F_{1,20} = 1.69$, $p = .21$) indicating that choices were similar

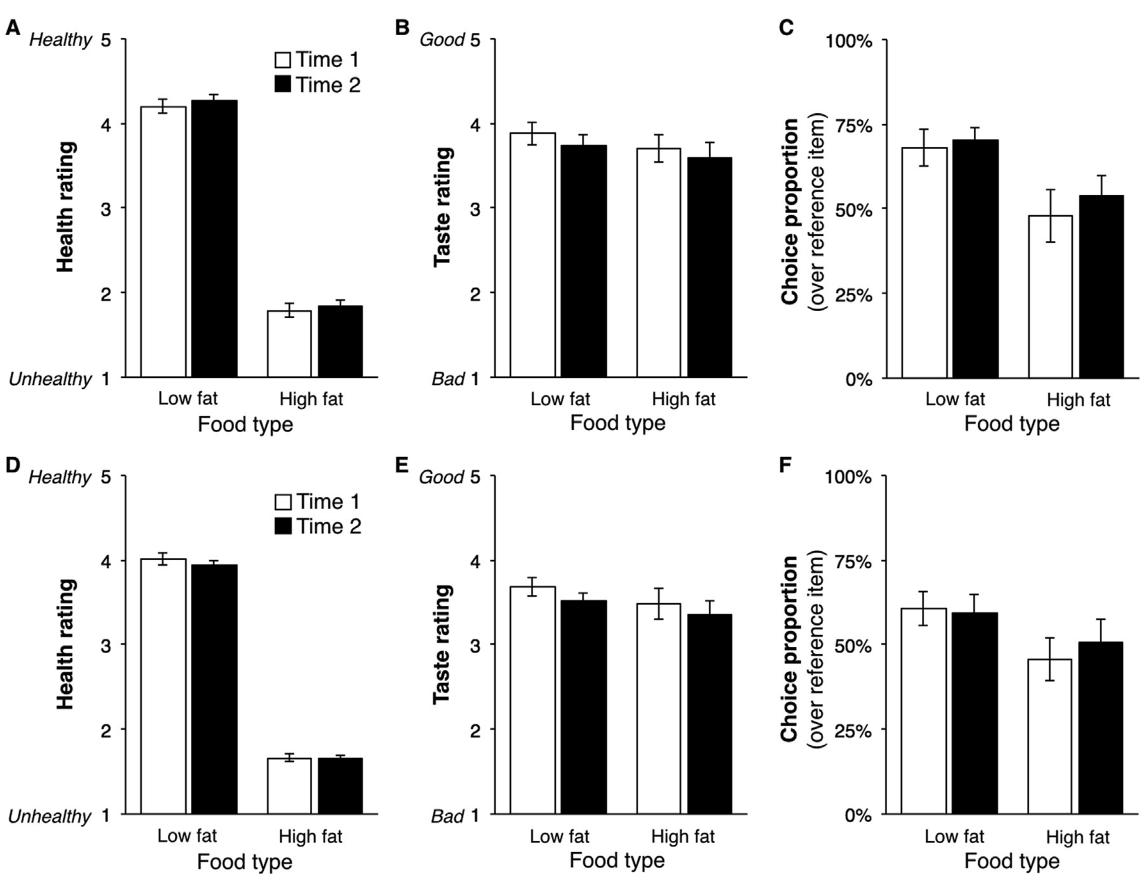

Fig. 2. Health rating (A), Taste rating (B), and Food choice (C) behavior in Study 1 showed no significant change over time. Health rating (D), Taste rating (E), and Food choice (F) behavior in Study 2 also showed no significant change over time.

across Time 1 and Time 2.

4. Discussion

Across two studies, the test-retest reliability of the Food Choice Task in healthy adult volunteers was very good, as evidenced by intra-class correlations. Reliability was high at both shorter (~3 days) and longer (~1 month) test-retest intervals. These results suggest that the Food Choice Task may be useful for measuring food-based decision-making in studies with multiple assessment points. For example, the stability of the outcomes over time in healthy volunteers suggests the task may be useful for evaluating the role of interventions that aim to alter eating behavior. Such interventions are relevant for a broad range of dysfunctional eating behavior.

A strength of this particular food-based decision task is that it allows for examination of restrictive food intake. On this task, restriction is quantified as the proportion of choices of high-fat foods over the individualized reference item because restriction commonly involves avoidance of fat. Test-retest reliability was particularly high for high-fat foods, and in Study 2 the reliability of ratings was significantly higher for choices in the high-fat trials than the low-fat trials. Therefore, the task may be particularly useful for assessing the level of restrictive eating, which can be challenging to measure because it is, in part, the absence of a behavior—the absence of eating. For example, the proportion of choices of high-fat foods has been shown to clearly distinguish between individuals with and without AN (Foerde et al., 2015), and among individuals with AN, choices of high-fat foods on the task were associated with actual caloric intake in a lunch meal, thereby linking task behavior with real eating behavior (Foerde et al., 2015). Restrictive eating behavior is a major contributor to the morbidity and mortality of individuals with AN, and therefore development of interventions that aim to increase food intake and dietary flexibility is critical. Additionally, the Food Choice Task may be useful for examining interventions for obesity and characterizing the food choices of individuals with obesity who are attempting to restrict caloric intake (Gianini, Walsh, Steinglass, & Mayer, 2017). The current study demonstrates that among healthy volunteers, who are not trying to alter their weight or eating behavior, performance on the task does not change substantially upon repeat administration, supporting the utility of this task as a tool to examine the impact of treatments and interventions on food choice among individuals who are attempting to change their food consumption.

We also examined the reliability of ratings of individual food items. Across both studies, reliability was “Moderate” or better for the majority of food items, but varied considerably across items. Surprisingly, the Health phase included more items with low reliability scores. Inspection of the data showed that this occurred when most or all participants provided identical ratings (for example, all participants rated peaches or green beans as very healthy). In the absence of any variability in the ratings, reliability is undefined as it measures agreement among ratings that vary across individuals. Although some items were rated as having only “Slight” or “Fair” reliability, in the aggregate task reliability was very good.

There are several potential limitations to consider in interpreting this study. One, we examined task reliability over a relatively short time frame and tested a single repetition. Although it is possible that the stability of ICCs could be limited by having only a single repetition, it is reassuring that we were able to demonstrate this effect across two independent studies. Two, whereas it is possible that the short time frame allows for carry-over effects or recall (Vaz, Falkmer, Passmore, Parsons, & Andreou, 2013), the absence of a learning component to the task diminishes the concern. Three, in Study 1, there were non-significant trends toward an increase in healthiness ratings and a decrease in tastiness ratings over time. However, the proportion of high-fat foods selected did not change over time. Thus any shifts in health and taste ratings were not associated with a change in the measure of restrictive food intake. Four, one of the two groups received an affect manipulation. However, we found that this did not affect health or taste ratings, or choice of high and low fat foods, and if it had, it would be likely to influence the results toward decreasing reliability. Finally, sample sizes for both studies were modest, albeit in the range of typical studies using patient populations. However, reliability was demonstrated in two studies, which bolsters confidence in the results. In summary, the Food Choice Task was found to have good test-retest reliability across two data sets, suggesting that it is well suited to longitudinal assessment of food choices.

Acknowledgements

Funding: This work was supported by the National Institutes of Health [grant numbers NIMH K23 MH076195, NIMH R01 MH079397], and the Global Foundation for Eating Disorders.

References

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290.

Efron, B., & Tibshirani, R. (1994). *An introduction to the bootstrap*. CRC press.

Fairburn, C. G., & Cooper, P. J. (1993). In C. G. Fairburn, & G. T. Wilson (Eds.), *Binge eating: Nature, assessment, and treatment* (pp. 317–360). Guilford Press.

Feng, Y., & Wen, V. (2010). *A program to automatically compute agreement Statistics for an asymmetric table*.

Foerde, K., Steinglass, J. E., Shohamy, D., & Walsh, B. T. (2015). Neural mechanisms supporting maladaptive food choices in anorexia nervosa. *Nature Neuroscience, 18*(11), 1571–1573.

Gianini, L., Riegel, M., Steinglass, J., Foerde, K., Attia, E., & Walsh, B. T. (2016). Relationship between negative affect and food choice in bulimia nervosa. In *Paper presented at the international conference on eating disorders* (San Francisco, CA).

Gianini, L. M., Walsh, B. T., Steinglass, J., & Mayer, L. (2017). Long-term weight loss maintenance in obesity: Possible insights from anorexia nervosa? *International Journal of Eating Disorders, 50*(4), 341–342.

Hadigan, C. M., Anderson, E. J., Miller, K. K., Hubbard, J. L., Herzog, D. B., Klibanski, A., et al. (2000). Assessment of macronutrient and micronutrient intake in women with anorexia nervosa. *International Journal of Eating Disorders, 28*(3), 284–292.

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science, 324*(5927), 646–648.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Mayer, L. E., Schebendach, J., Bodell, L. P., Shingleton, R. M., & Walsh, B. T. (2012). Eating behavior in anorexia nervosa: Before and after treatment. *International Journal of Eating Disorders, 45*(2), 290–293.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Current Opinion in Neurobiology, 14*(6), 769–776.

Rangel, A. (2013). Regulation of dietary choice by the decision-making circuitry. *Nature Neuroscience, 16*(12), 1717–1724.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.

Spitzer, R. L., Williams, J. B. W., & Gibbon, M. (1987). *Structured clinical interview for DSM-IV-R (SCID)*. New York: New York State Psychiatric Institute, Biometrics Research.

Steinglass, J., Foerde, K., Kostro, K., Shohamy, D., & Walsh, B. T. (2015). Restrictive food intake as a choice-A paradigm for study. *International Journal of Eating Disorders, 48*(1), 59–66.

Team, R Core. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One, 8*(9), e73990.

Walsh, B. T. (2011). The importance of eating behavior in eating disorders. *Physiology & Behavior, 104*(4), 525–529.

Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). Guidelines for estimating repeatability. *Methods in Ecology and Evolution, 3*(1), 129–137.